



ELSEVIER



# Convergent and divergent mechanisms of sugar recognition across kingdoms

Maureen E Taylor and Kurt Drickamer

Protein modules that bind specific oligosaccharides are found across all kingdoms of life from single-celled organisms to man. Different, overlapping and evolving designations for sugar-binding domains in proteins can sometimes obscure common features that often reflect convergent solutions to the problem of distinguishing sugars with closely similar structures and binding them with sufficient affinity to achieve biologically meaningful results. Structural and functional analysis has revealed striking parallels between protein domains with widely different structures and evolutionary histories that employ common solutions to the sugar recognition problem. Recent studies also demonstrate that domains descended from common ancestors through divergent evolution appear more widely across the kingdoms of life than had previously been recognized.

## Addresses

Department of Life Sciences, Imperial College, London SW7 2AZ, United Kingdom

Corresponding author: Drickamer, Kurt ([k.drickamer@imperial.ac.uk](mailto:k.drickamer@imperial.ac.uk))

Current Opinion in Structural Biology 2014, 28:14–22

This review comes from a themed issue on **Carbohydrate-protein interactions and glycosylation**

Edited by **Harry J Gilbert** and **Harry Brumer**

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online 5th August 2014

<http://dx.doi.org/10.1016/j.sbi.2014.07.003>

0959-440X/© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

## Introduction

Proteins that seem to have a primary function of binding sugars are often referred to as lectins, a term used initially in the context of plant seed proteins and then broadened to include examples from a wider range of species [1]. However, the names of many proteins that have sugar-binding activity are based on other biological functions that they have. For example, plant toxins represent a group of proteins in which sugar-binding activity in one part of a protein is used to target killing functions of another part of the protein. Similarly, sugar-binding proteins in yeast are usually denoted by their functions in flocculation and in adhesion. Bacterial proteins that interact with oligosaccharide ligands include adhesins, on fimbriae and pili [2], as well as toxins, but there is also

sugar-binding activity associated with many glycosidases that contain non-catalytic carbohydrate-binding modules [3]. It is also common to use alternative designations such as glycan-binding proteins or glycan-binding receptors, particularly in the case of animal lectins.

In spite of this diversity of names and functions, a common feature of all of these proteins is that the sugar recognition function in each protein is mediated by a discrete protein module. The term carbohydrate-recognition domain is often used as a general label that encompasses all of the diverse folds, functions and sites of expression. However, many of the individual groups described in this review have other, more common designations and no systematic revision of the nomenclature seems appropriate at this point. Nevertheless, it is important that the diversity of names and categories does not obscure many evolutionary relationships between carbohydrate-recognition proteins, domains and modules in different species and kingdoms of life arising through divergent evolution as well as interesting similarities in the mechanisms of carbohydrate recognition that have come about through convergent evolution.

## Carbohydrate recognition in multiple protein fold families

One approach to comparing mechanisms of sugar recognition is to classify carbohydrate-recognition domains based on their sequences and structures. Two key conclusions emerge from such comparisons and the resulting classifications. First, sugar-binding activity can appear in the context of many different protein folds. Second, the protein folds of carbohydrate-recognition domains are not exclusively associated with sugar-binding activity. The first of these conclusions reflects independent evolution of this activity on multiple occasions and means that there is no simple way to identify sugar-binding proteins by looking for one particular protein fold [4]. The second conclusion means that similarity in the fold of a novel domain to a fold that can support sugar-binding activity does not necessarily imply that the new domain will bind sugars.

The principle that fold does not necessarily imply function is well established in the case of the C-type carbohydrate-recognition domains of animal lectins, which are a subset of the broader family designated C-type lectin-like domains that includes many members that lack sugar-binding activity [5], including some for which reports of sugar-binding activity have recently been called

into question [6]. In many cases, other target ligands, such as lipoproteins or other proteins, are known but in other instances the functions of these domains remain to be established. Similar principles are evident for mannose 6-phosphate receptor homology domains (MRH domains), only some of which bind sugars, while others have ligands such as insulin-like growth factor II [7]. The same ideas emerge again for Fbs proteins that target tagging of proteins with ubiquitin by binding to the chitobiose core of N-linked glycans [8]. The Fbs proteins are part of a larger family of F-box proteins, most of which do not bind sugars and in fact at least one Fbs protein appears to lack this activity [9].

### Convergence on shared features in monosaccharide-specific sites

A further consequence of these insights is that common features in the mechanisms of recognition of sugars that transcend fold families reflect convergent evolution. Two such features that crop up remarkably often are packing of sugars against aromatic residues and involvement of  $\text{Ca}^{2+}$ . The former type of interaction, particularly between the apolar B face of galactose and a tryptophan residue, has been extensively discussed [10]. Ligation of sugars to  $\text{Ca}^{2+}$  ions was first described for the C-type carbohydrate-recognition domains in animal lectins [11], but has recently been identified in several other groups of sugar-binding proteins with carbohydrate-recognition domains from different fold families (Figure 1). Examples include yeast flocculation proteins [12] and adhesins [13<sup>••</sup>,14<sup>••</sup>] and at least two families of bacterial carbohydrate-binding modules [15], as well as the processing mannoside from the endoplasmic reticulum [16]. Other sugar-binding proteins that employ a pair of  $\text{Ca}^{2+}$  in sugar-binding sites are the pentraxin serum amyloid protein [17] and the lectin from *Pseudomonas aeruginosa* [18]. The convergent use of  $\text{Ca}^{2+}$  ligation in different structural contexts reflects the fundamental chemistry of the sugars, which are known to bind free  $\text{Ca}^{2+}$  [19].

In contrast to the cases noted above,  $\text{Ca}^{2+}$  and other divalent cations are sometimes indirectly involved in sugar binding, because they stabilize the sugar-binding conformation of a carbohydrate-recognition domain, for example in legume lectins [20], calnexin and calreticulin [21,22] and at least one family of bacterial carbohydrate-binding modules [23]. An interesting recent example of such an arrangement is seen in the mammalian L-type lectins ERGIC-53 and VIP36. On the basis of recent structural analysis of ERGIC-53, also known as LMAN1, it has been suggested that modulation of binding by different  $\text{Ca}^{2+}$  concentrations occurs in various luminal compartments in cells [24<sup>••</sup>]. This phenomenon appears to be a more subtle form of modulation of sugar-binding activity than that observed for the endocytic C-type lectins, in which loss of  $\text{Ca}^{2+}$  binding at endosomal pH results in loss of sugar binding activity, which provides a

means of separating endocytic cargo from the receptors [25,26].

### A proliferation of secondary binding sites

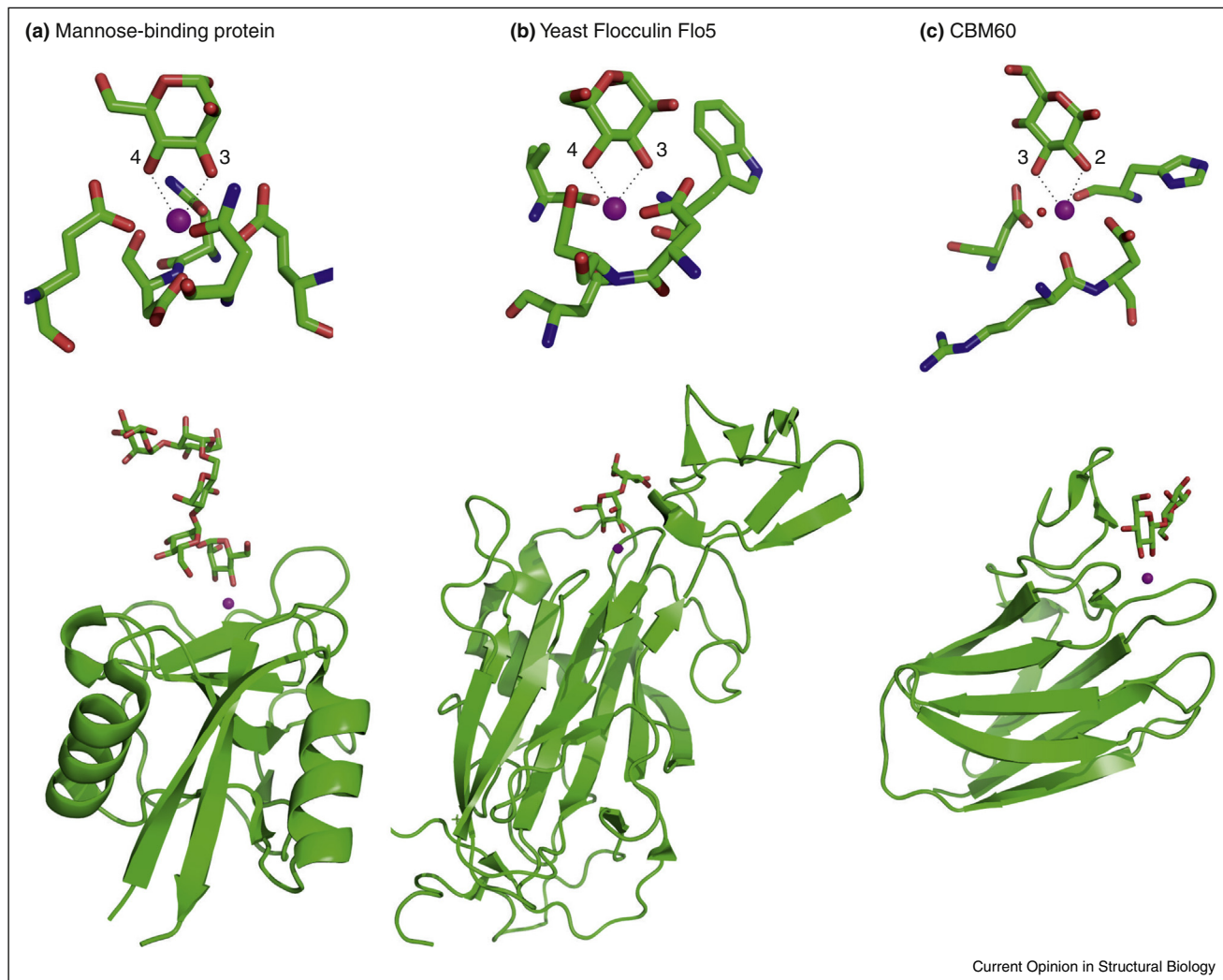
A further interesting comparison of convergent sugar-binding sites is that, within fold families, there are often common mechanisms of binding to a core monosaccharide in a primary binding site, but diversity in binding of oligosaccharide and glycoconjugate ligands is achieved through extended and secondary binding sites that are unique to individual members of the family. Such extensions can involve interactions with additional sugar residues in an oligosaccharide ligand, but an increasing number of examples demonstrate binding to other modifications of the sugars.

Differences in secondary or extended binding sites often provide specificity for different oligosaccharides in closely related proteins. For example, the sorting lectins ERGIC-53 and VIP36 bind to distinct groups of high mannose oligosaccharides using a common primary mannose binding site that is extended in different ways. Because of these differences, ERGIC-53 binds to any  $\text{Man}\alpha 1\text{-}2\text{Man}$  disaccharides, including one bearing  $\text{Glc}\alpha 1\text{-}3$  substitution on the non-reducing terminus [24<sup>••</sup>] compared to the selectivity of VIP36 for the unglucosylated  $\text{Man}\alpha 1\text{-}2\text{Man}\alpha 1\text{-}2\text{Man}\alpha 1\text{-}3$  arm of high mannose oligosaccharides [27].

In the C-type lectin family, multiple examples of secondary interactions with sugars are common, leading to binding of high mannose and Lewis<sup>x</sup> motifs, for example [28] and non-sugar substituents such as sulfate can also be accommodated in secondary binding sites [29]. In a novel twist, the macrophage receptor mincle has recently been shown to bind trehalose, the glucose  $\alpha 1\text{-}1$  glucose disaccharide, through such a mechanism, but further specificity for mycobacterial glycolipids that bear this headgroup is achieved through interactions with the attached hydrophobic acyl chains, apparently through an adjacent hydrophobic groove [30<sup>••</sup>] (Figure 2).

In the case of the mannose 6-phosphate receptors, a mechanism involving a mannose-binding site extended by secondary interactions with a phosphate substituent is well established [31]. Elaboration of the secondary binding site, leading to selectivity for a  $\text{GlcNAc}$  residue attached to mannose through a phosphodiester linkage, can be achieved by a combination of removing potential hindrance to binding of the  $\text{GlcNAc}$  residues with addition of favourable secondary interactions between the protein and the added sugar [32]. The MRH domain in the OS-9 protein, which forms part of the quality control system of the endoplasmic reticulum, provides an alternative variation on the binding site in which a pair of tryptophan residues extends the primary mannose-binding site, making it selective for oligosaccharides containing  $\text{Man}\alpha 1\text{-}6\text{Man}$  units [33]. In contrast, recent

Figure 1



Involvement of Ca<sup>2+</sup> in sugar-binding sites in the context of multiple different protein folds. Top panels show close-up views of sugar-binding sites and lower panels show overall folds of carbohydrate-recognition domains. **(a)** Human serum mannose-binding protein (2MSB) with Man<sub>5</sub> oligosaccharide. **(b)** Yeast flocculin Flo5 (2XJS) with bound mannobiose. **(c)** Family 60 carbohydrate-binding module (CBM60) from *Cellvibrio japonicus* xylanase (2XFD) with bound cellobiose. Ca<sup>2+</sup> is indicated as a magenta sphere in each panel and water is represented as a red sphere. Coordination bonds from adjacent equatorial hydroxyl groups to the Ca<sup>2+</sup> are indicated as dashed lines.

analysis of the MRH domain in endoplasmic reticulum glucosidase II reveals an open binding site that lacks any of these extensions and thus represents a more prototypical mannose-binding site [34<sup>••</sup>].

### Common approaches to combining binding sites

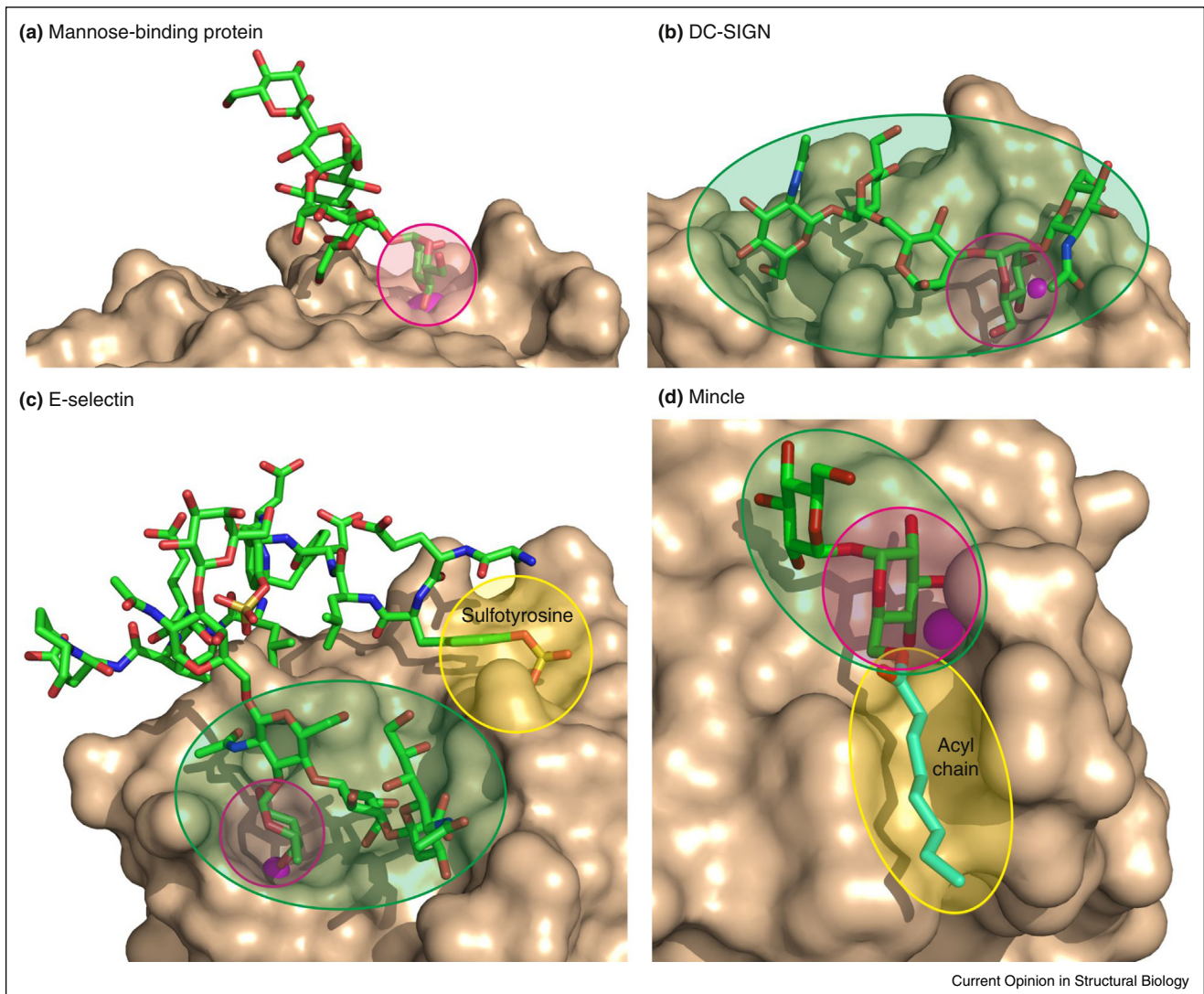
In addition to convergence in the way that binding sites in individual domains work, the arrangement of these domains within proteins shows some interesting parallels between different groups of sugar-binding proteins. The phenomenon of enhanced binding to multivalent ligands through clustering of binding sites in oligomers is well

established and has been extensively investigated for many types of sugar-binding proteins [35,36].

Somewhat less appreciated has been the generality of an arrangement in which a carbohydrate-recognition domain targets and enhances the activity of an enzyme that builds or degrades carbohydrates (Figure 3). The most extensively studied examples of such an arrangement are the carbohydrate-binding modules linked to the catalytic domains of many polysaccharide hydrolases [37<sup>•</sup>]. The recent demonstration of how an MRH domain linked to the catalytic domain of endoplasmic reticulum glucosidase II enhances the activity of this enzyme on nascent N-linked glycans demonstrates that similar pairings of



Figure 2



Multiple different ways in which binding specificity of C-type carbohydrate-recognition domains is enhanced by extended and accessory binding sites. Each of the binding sites involves a primary interaction between the  $\text{Ca}^{2+}$ , shown in magenta, and two adjacent hydroxyl groups on a monosaccharide residue. **(a)** The relatively open binding site in mannose-binding protein binds only a terminal mannose residue, so only this residue interacts with the protein (2MSB). **(b)** DC-SIGN binds a more complex  $\text{Man}_3\text{GlcNAc}_2$  oligosaccharide through an extended binding site that accommodates sugars on either side of the mannose residue in the primary binding site (1K9J). **(c)** In addition to ligation of fucose to  $\text{Ca}^{2+}$ , the sialyl Lewis<sup>x</sup> oligosaccharide interacts with an extended binding site in E-selectin, which also has an accessory binding site for sulfated tyrosine residues on a glycoprotein ligand (1G1S). **(d)** Mincle binds to the disaccharide trehalose as a result of one glucose residue binding in the  $\text{Ca}^{2+}$  site and the second glucose residue contacting an adjacent site. In addition, glycolipid binding is enhanced through an accessory site that forms a hydrophobic groove which can interact with acyl chains on the 6-OH groups of the glucose residues (4KZV). Primary binding sites are highlighted in pink, extended oligosaccharide-binding sites are indicated in green and accessory sites for other modifications are shaded yellow.

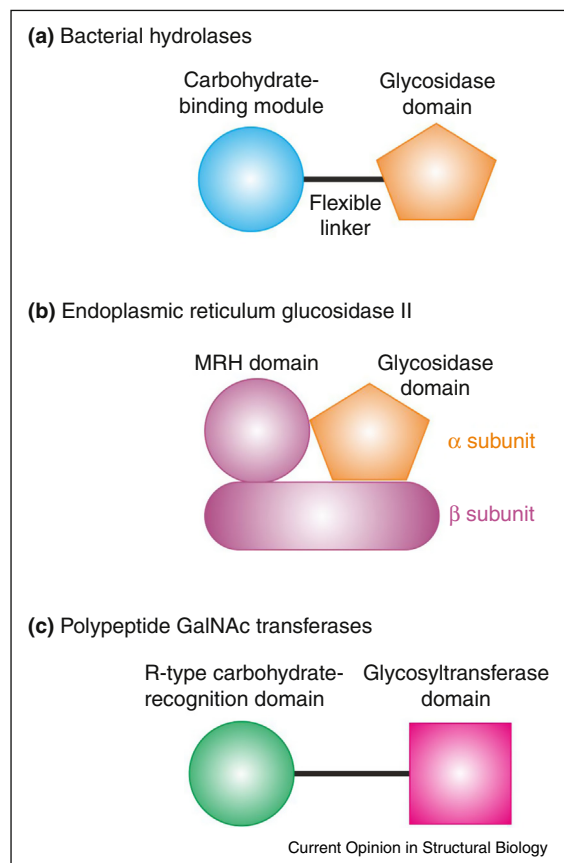
sugar-binding and catalytic domains can be achieved using completely different structural elements [34<sup>\*\*</sup>]. The same principle is seen in the large family of polypeptide *N*-acetylgalactosaminyltransferases, but in these cases it is R-type carbohydrate recognition domains coupled to synthetic enzymes that target the enzymes to sites adjacent to already glycosylated residues [38,39<sup>\*</sup>]. Recent examples also illustrate how a carbohydrate-binding module can effectively extend the active site of a

hydrolase [40] and that PA14 carbohydrate-binding modules can be inserted into the hydrolase domains rather than just being appended to them [41,42].

### Some old distinctions becoming less clear

It is increasingly difficult to delineate well defined subgroups of sugar-binding proteins based on any features other than sequence similarity. For example, as noted in the previous section, a domain organization linking a

Figure 3



Association of carbohydrate-recognition domains with enzymatically active domains. **(a)** One or more carbohydrate-binding modules are often linked to bacterial glycosidases and cellulose-degrading enzymes in a single polypeptide. The carbohydrate-binding modules localize the activity on substrates and enhance the activity of enzymes. **(b)** The  $\alpha$  subunit of endoplasmic reticulum glucosidase II contains the glucosidase active site, but the activity of the enzyme on high mannose oligosaccharides that bear terminal glucose residues on one branch is enhanced by the  $\beta$  subunit, which contains an MRH domain that binds mannose on another branch of the oligosaccharide. **(c)** R-type carbohydrate-recognition domains in many of the polypeptide GalNAc transferase proteins direct the enzyme to regions of substrate glycoproteins that already bear one or more GalNAc residues.

domain that recognizes sugars with one that catalyses modification of the sugar is no longer just a feature of the carbohydrate-binding module/glycosidase family. At the same time, some of the carbohydrate-binding modules linked to hydrolase domains are structurally related to lectins that are separate from catalytic domains [43]. Thus, a particular domain organization is not uniquely associated with a particular structural group of carbohydrate-recognition domains. Similarly, while hydrolase-associated carbohydrate-binding modules are often associated with binding of internal sugars in polysaccharide chains, a significant subgroup of these domains are now known to bind non-reducing terminal residues [37].

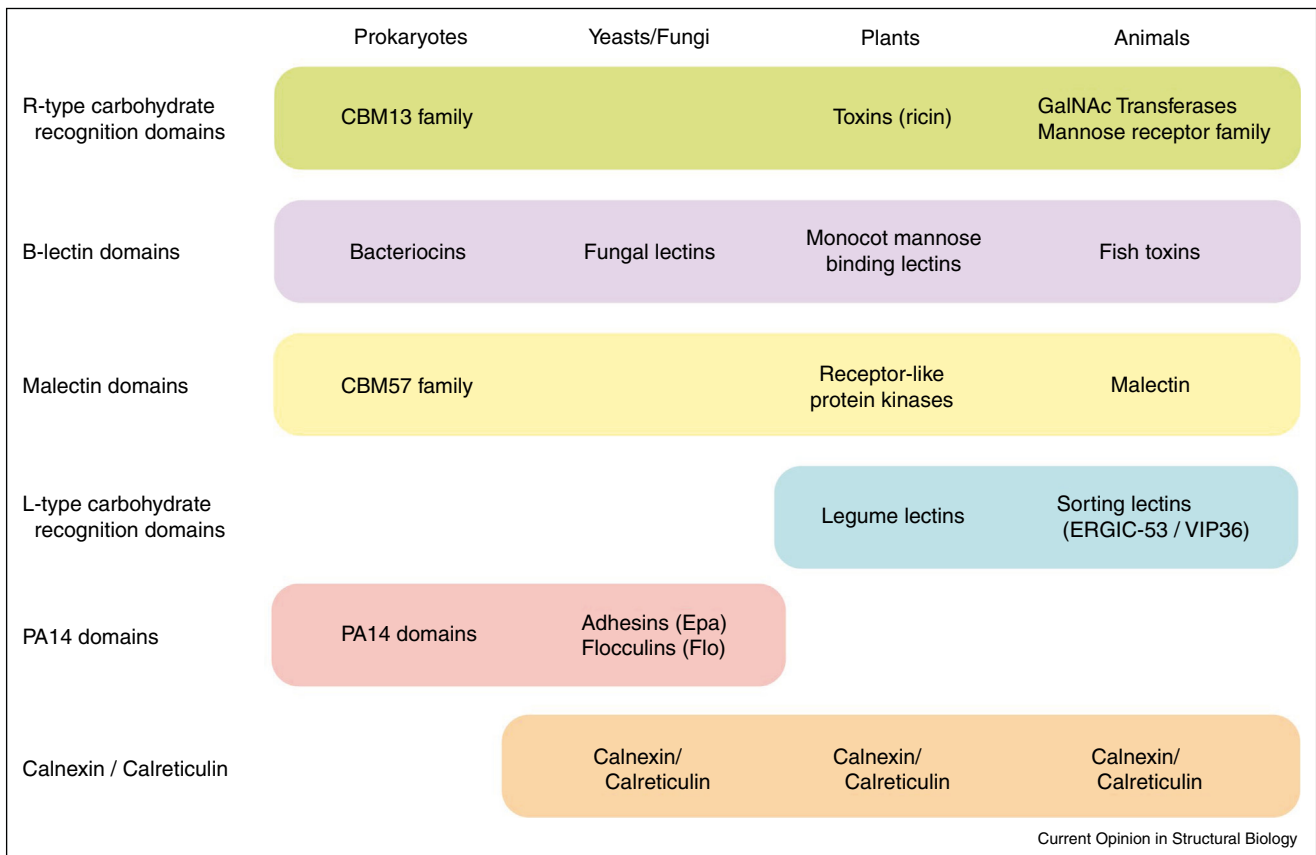
Conversely, not all proteins referred to as lectins bind terminal residues, since the ability of galectins to interact with residues within a polypeptide is now well established [44].

Perhaps the most interesting recent change in the perception of different groups of carbohydrate-recognition domains is the finding that many of the families, for which sequence similarity provides strong evidence of divergence from a common ancestor, appear in a more diverse range of species and even kingdoms of life than was previously appreciated (Figure 4). It was previously recognized that structurally related domains used for different functions appear across the animal and plant kingdoms, since L-type carbohydrate-recognition domains are found in the legume lectins in plants as well as the sorting lectins such as ERGIC-53 and VIP36 in animal cells [24,27]. It is now clear that structurally related carbohydrate-binding domains are present in both eukaryotes and prokaryotes. One of the most widely spread type of domain is the R-type carbohydrate-recognition domain, originally described in plant toxins such as ricin [45] and more recently recognized in polypeptide N-acetylgalactosaminyltransferases [39] and the mannose-receptor family of proteins in animals [46] as well as the bacterial glycoside hydrolases containing CBM13 modules [47]. A second widely represented family of domains is the monocot mannose B-lectin type domain, widely studied in plants but also described in fish and fungi [48,49] and more recently in bacteria, including bacteriocins from *Pseudomonas* [50,51]. A third family that spans from prokaryotes to eukaryotes is the PA14 domain, which exhibits carbohydrate binding activity both as a carbohydrate-binding module in bacterial glycosidases and in yeast adhesions and flocculation factors [52]. A further unexpected sequence relationship is that between the endoplasmic reticulum sorting lectin mallectin [53] and the CBM57 family of carbohydrate-binding modules of bacterial glycosidases and similar domains in putative plant kinases [54], although in the latter case the apparent sequence similarities remain to be followed up with evidence for structural similarity and sugar-binding activity. These observations reflect the role of divergence of sugar binding domains as well as importance of convergence on similar recognition principles.

### Polymorphism analysis

A number of interesting patterns have been observed in the evolution of several of the families of glycan-binding receptors. Within the mammalian families, some types of receptors, such as those involved in intracellular trafficking of glycoproteins, are often relatively conserved across species, but some of the cell surface receptors tend to be more divergent. Extreme examples of such divergence include the DC-SIGN homologs [55] and the CD33-related siglecs [56]. Evolution of variability in receptors of the innate immune system probably reflects selective

Figure 4



A summary of some of the types of carbohydrate-recognition domains that are found in a wide range of species. Other types of domain not shown are expressed only in more restricted groups of organisms. For example, galectins, siglecs and C-type lectins are expressed only in animals and several classes of adhesins are specific to bacteria.

pressure from rapidly changing pathogens, many of which can exploit glycan-binding receptors as a means of entering cells.

In addition to variation between species, selection pressure from pathogens has led to establishment of polymorphisms in some of the receptors. The best studied example of a balanced polymorphism is in serum mannose-binding protein, in which several variants that have reduced capacity to activate complement have been identified [57]. The structural basis for how changes in the structure of the collagen-like domains in mannose-binding protein affect the interaction with complement is at least partially understood [58]. There is strong genetic evidence that other polymorphisms that result in amino acid substitutions in glycan-binding receptors of the innate immune system can affect susceptibility to infection. For example, polymorphisms in the C-type carbohydrate-recognition domains of the mannose receptor are linked to susceptibility to leprosy [59] and variability in the number of repeats in the neck region of DC-SIGNR, the endothelial paralog of DC-SIGN, may

impact on transmission of human immunodeficiency virus [60]. However, in these latter cases, the molecular mechanisms that underlie the phenotypic consequences of changes to the amino acids sequences of these proteins remain to be established.

Other variations in the sequences of glycan-binding receptors have been more directly linked to changes in the sugar-binding activity of these proteins. Recent studies on langerin reveal that a form of the protein containing two amino acid changes compared to the most common reference form undergoes a major change in ligand binding, in which the ability to bind glycans terminated with galactose 6-sulphate is lost, while the affinity for glycans terminating in *N*-acetylglucosamine is increased [61<sup>••</sup>]. In this case, the amino acid changes are directly in the binding site and a structural basis for the changes in sugar binding has been demonstrated. The langerin results provide a paradigm for a novel way in which the diversity of glycan-binding receptors can be increased. In contrast, although there is increasing genetic evidence for linkage of risk of coronary artery disease with



variants in the epidermal growth factor domain adjacent to the C-type carbohydrate-recognition domain of E-selectin [62], recent attempts to verify previous suggestions that these changes alter the sugar-binding activity of the receptor have been unsuccessful [63\*]. From a structural perspective, this outcome is probably not surprising, given that the polymorphism is distant from the ligand-binding site and in a separate domain.

## Conclusions

It is clear that there is no single set of unifying principles that describe carbohydrate recognition across all the kingdoms of life. Nevertheless, the examples described in this short review illustrate that some of the solutions to the sugar recognition problem go back very far in evolution and that mechanisms for binding sugars based on the chemical properties of the sugar ligands can be implemented in the context of many different protein folds. The first of these conclusions provides a useful basis for identifying potential sugar recognition systems from genomic sequence data. However, the second point means that novel carbohydrate-recognition domains which utilize different protein scaffolds may still remain to be discovered.

## Acknowledgements

This work was supported by grant BB/K007718/1 from the Biotechnology and Biological Sciences Research Council and grant 093599 from the Wellcome Trust.

## References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Sharon N, Lis H: **History of lectins: from hemagglutinins to biological recognition molecules.** *Glycobiology* 2004, **14**:53R-62R.
  2. Etzold S, Juge N: **Structural insights into bacterial recognition of intestinal mucins.** *Curr Opin Struct Biol* 2014, **28** (in press).
  3. Abbott WD, van Bueren AL: **Using structure to inform carbohydrate binding module function.** *Curr Opin Struct Biol* 2014, **28** (in press).
  4. Drickamer K, Taylor ME: **Identification of lectins from genomic sequence data.** *Methods Enzymol* 2003, **362**:592-599.
  5. Weis WI, Taylor ME, Drickamer K: **The C-type lectin superfamily in the immune system.** *Immunity* 1998, **163**:19-34.
  6. Rozbeský D, Krejzová J, Křenek K, Prchal J, Hrabal R, Kožíšek M, Weignerová L, Fiore M, Dumy P, Křen V *et al.*: **Re-evaluation of binding properties of recombinant lymphocyte receptors NKR-P1A and CD69 to chemically synthesized glycans and peptides.** *Int J Mol Sci* 2014, **15**:1271-1283.
  7. Kim J-JP, Olson LJ, Dahms NM: **Carbohydrate recognition by the mannose-6-phosphate receptors.** *Curr Opin Struct Biol* 2009, **19**:534-542.
  8. Yoshida Y, Tokunaga F, Chiba T, Iwai K, Tanaka K, Tai T: **Fbs2 is a new member of the E3 ubiquitin ligase family that recognizes sugar chains.** *J Biol Chem* 2003, **278**:43877-43884.
  9. Glenn KA, Nelson RF, Wen HM, Mallinger AJ, Paulson HL: **Diversity in tissue expression, substrate binding, and SCF complex formation for a lectin family of ubiquitin ligases.** *J Biol Chem* 2008, **283**:12717-12729.
  10. Vandenbussche S, Díaz D, Fernández-Alonso MC, Pan W, Vincent SP, Cuevas G, Cañada FJ, Jiménez-Barbero J, Bartik K: **Aromatic-carbohydrate interactions: an NMR and computational study of model systems.** *Chem Eur J* 2008, **14**:7570-7578.
  11. Weis WI, Drickamer K, Hendrickson WA: **Structure of a C-type mannose-binding protein complexed with an oligosaccharide.** *Nature* 1992, **360**:127-134.
  12. Veelders M, Brückner S, Ott D, Unverzagt C, Mösch H-U, Essen L-O: **Structural basis of flocculin-mediate social behavior in yeast.** *Proc Natl Acad Sci U S A* 2010, **107**:22511-22516.
  13. Maestre-Reyna M, Diderrich R, Veelders M, Eulenburg G, Kalugin V, Brückner S, Keller P, Rupp S, Mösch H-U, Essen L-O: **Structural basis for promiscuity and specificity during *Candida glabrata* invasion of host epithelia.** *Proc Natl Acad Sci U S A* 2012, **109**:16864-16869.
- Two structures of yeast adhesins reveal that they resemble the flocculins in fold and mechanism of sugar binding, although their primary interaction is with galactose rather than mannose.
14. Ielasi FS, Decanniere K, Willaert RG: **The epithelial ashesin 1 (Epa1p) from the human-pathogenic yeast *Candida glabrata*: structural and functional study of the carbohydrate-binding domain.** *Acta Crystallogr* 2012, **D68**:210-217.
- See annotation to Ref [13\*\*].
15. Montanier C, Flint JE, Xie NBDH, Liu Z, Rogowski A, Weiner DP, Ratnaparkhe S, Nurizzo D, Roberts SM *et al.*: **Circular permutation provides an evolutionary link between two families of calcium-dependent carbohydrate binding modules.** *J Biol Chem* 2010, **285**:31742-31754.
  16. Vallée F, Karaveg K, Herscovics A, Moremen KW, Howell PL: **Structural basis for catalysis and inhibition of N-glycan processing class I  $\alpha$ 1,2-mannosidases.** *J Biol Chem* 2000, **275**:41287-41298.
  17. Thompson D, Pepys MB, Tickle I, Wood S: **The structures of crystalline complexes of human serum amyloid P component with its carbohydrate ligand, the cyclic pyruvate acetal of galactose.** *J Mol Biol* 2003, **320**:1081-1086.
  18. Perret S, Sabin C, Dumon C, Pokorná M, Gautier C, Galanina O, Iliá S, Bovin N, Nicaise M, Desmadril M *et al.*: **Structural basis for the interaction between human milk oligosaccharides and the bacterial lectin PA-IIL of *Pseudomonas aeruginosa*.** *Biochem J* 2005, **389**:325-332.
  19. Cook WJ, Bugg CE: **Calcium binding to galactose: crystal structure of a hydrated  $\alpha$ -galactose-calcium bromide complex.** *J Am Chem Soc* 1973, **95**:6442-6446.
  20. Weis WI, Drickamer K: **Structural basis of lectin-carbohydrate interaction.** *Annu Rev Biochem* 1996, **65**:441-473.
  21. Schrag JD, Bergeron JJM, Li Y, Borisova S, Hahn M, Thomas DY, Cygler M: **The structure of calnexin, an ER chaperone involved in quality control of protein folding.** *Mol Cell* 2001, **8**:633-644.
  22. Kozlov G, Pocanschi CL, Rosenauer A, Bastos-Aristizabal S, Gorelik A, Williams DB, Gehring K: **Structural basis of carbohydrate recognition by calreticulin.** *J Biol Chem* 2010, **285**:38612-38620.
  23. Henshaw J, Horne-Bitschy A, van Bueren AL, Money VA, Bolam DN, Czjzek M, Ekborg NA, Weiner RM, Hutcheson SW, Davies GJ *et al.*: **Family 6 carbohydrate binding modules in beta-agarases display exquisite selectivity for the non-reducing termini of agarose chains.** *J Biol Chem* 2006, **281**:17099-17107.
  24. Zheng C, Page RC, Das V, Nix JC, Wigren E, Misra S, Zhang B: **Structural characterization of carbohydrate binding by LMAN1 protein provides new insight into endoplasmic reticulum export of factors V and (FV) and VIII (FVIII).** *J Biol Chem* 2013, **288**:20499-20509.
- In addition to providing a basis for understanding differences in oligo-saccharide-binding specificity of L-type lectins, this manuscript suggests how their different trafficking properties may reflect modulation of sugar

binding by Ca<sup>2+</sup> concentrations in different luminal compartments.

25. Wragg S, Drickamer K: **Identification of amino acid residues that determine pH-sensitivity of ligand binding to the asialoglycoprotein receptor.** *J Biol Chem* 1999, **274**:35400-35406.
26. Feinberg H, Torgersen D, Drickamer K, Weis WI: **Mechanism of pH-dependent N-acetylgalactosamine binding to a functional mimic of the hepatic asialoglycoprotein receptor.** *J Biol Chem* 2000, **275**:35176-35184.
27. Satoh T, Cowieson NP, Hakamata W, Ideo H, Fukushima K, Kurihara M, Kato R, Yamashita K, Watatsuki S: **Structural basis for recognition of high mannose type glycoproteins by mammalian transport lectin VIP36.** *J Biol Chem* 2007, **282**:28246-28255.
28. Guo Y, Feinberg H, Conroy E, Mitchell DA, Alvarez R, Blixt O, Taylor ME, Weis WI, Drickamer K: **Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR.** *Nat Struct Mol Biol* 2004, **11**:591-598.
29. Somers WS, Tang J, Shaw GD, Camphausent RT: **Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SleX and PSGL-1.** *Cell* 2000, **103**:467-479.
30. Feinberg H, Jégouzo SAF, Rowntree TJW, Guan Y, Brash MA, Taylor ME, Weis WI, Drickamer K: **Mechanism for recognition of an unusual mycobacterial glycolipid by the macrophage receptor mincle.** *J Biol Chem* 2013, **288**:28457-28465.

A novel accessory binding site in mincle that accommodates acyl chains provides a mechanism for enhanced affinity of this macrophage receptor for glycolipids.

31. Castonguay AC, Olson LJ, Dahms NM: **Mannose 6-phosphate receptor homology (MRH) domain-containing lectins in the secretory pathway.** *Biochim Biophys Acta* 2011, **1810**:815-826.
32. Olson LJ, Peterson FC, Castonguay AC, Bohnsack RN, Kudo M, Gotschall RR, Canfield WM, Volkman BF, Dahms NM: **Structural basis for recognition of phosphodiester-containing lysosomal enzymes by the cation-independent mannose 6-phosphate receptor.** *Proc Natl Acad Sci U S A* 2010, **107**:12493-12498.
33. Satoh T, Chen Y, Hu D, Hanashima S, Yamamoto K, Yamaguchi Y: **Structural basis for oligosaccharide recognition of misfolded glycoproteins by OS-9 in ER-associated degradation.** *Mol Cell* 2010, **40**:905-916.
34. Olson LJ, Orsi R, Alculumbre SG, Peterson FC, Stigliano ID, Parodi AJ, D'Alessio C, Dahms NM: **Structure of the lectin mannose 6-phosphate receptor homology (MRH) domain of glycosidase II, an enzyme that regulates glycoprotein folding quality control in the endoplasmic reticulum.** *J Biol Chem* 2013, **288**:16460-16475.

The structure of the MRH domain from glucosidase II provides a model for a primordial mannose-binding site as well as suggesting how this domain enhances the glycosidase activity of the catalytic subunit.

35. Lee RT, Lee YC: **Affinity enhancement by multivalent lectin-carbohydrate interaction.** *Glycoconj J* 2000, **17**:543-551.
36. Dam TK, Gerken TA, Brewer CF: **Thermodynamics of multivalent carbohydrate-lectin crosslinking interactions: importance of entropy in the bind and jump mechanism.** *Biochemistry* 2009, **48**:3822-3827.
37. Gilbert HJ, Knox JP, Boraston AB: **Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules.** *Curr Opin Struct Biol* 2013:669-677.

A recent summary provides a suggested reclassification of carbohydrate-binding modules from bacterial polysaccharide-degrading enzymes.

38. Hassan H, Reis CA, Bennett EP, Mirgorodskaya E, Roepstorff P, Hollingsworth MA, Burchell J, Taylor-Papadimitiou J, Clausen H: **The lectin domain of UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-T4 directs its glycopeptide specificities.** *J Biol Chem* 2000, **275**:38197-38205.
39. Gerken TA, Revoredo L, Thome JJ, Tabak LA, Vester-Christensen MB, Clausen H, Gahlay GK, Jarvis DL, Johnson RW, Moniz HA *et al.*: **The lectin domain of the polypeptide GalNAc**

**transferase family of glycosyltransferases (ppGalNAc Ts) acts as a switch directing glycopeptide substrate glycosylation in an N- or C-terminal direction, further controlling mucin type O-glycosylation.** *J Biol Chem* 2013, **288**:19900-19914.

Systematic analysis using glycopeptide substrates provides direct evidence for the effect of pre-existing GalNAc residues on the activity of the polypeptide GalNAc transferases.

40. Meekins MA, Raththagala M, Husodo S, White CJ, Guo H-F, Kötting O, van der Kooi CW, Gentry MS: **Phosphoglucan-bound structure of starch phosphatase Starch Excess4 reveals the mechanism for C6 specificity.** *Proc Natl Acad Sci U S A* 2014, **111**:7272-7277.
41. Yoshida E, Hidaka M, Fushinobu S, Koyanagi T, Minami H, Tamaki H, Kitaoka M, Katayama T, Kumagai H: **Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3  $\beta$ -glucosidase from *Kluyveromyces marxianus*.** *Biochem J* 2010, **431**:39-49.
42. Larsbrink J, Izumi I, Ibatullin FM, Nakhai A, Gilbert HJ, Davies GJ, Brumer H: **Structural and enzymatic characterization of a glycoside hydrolase family 31  $\alpha$ -xylosidase from *Cellvibrio japonicus* involved in xyloglucan saccharification.** *Biochem J* 2011, **436**:567-580.
43. Boraston AB, Notenboom V, Warren RAJ, Kilburn DG, Rose DR, Davies G: **Structure and ligand binding of carbohydrate-binding module CsCBM6-3 reveals similarities with fucose-specific lectins and galactose-binding domains.** *J Mol Biol* 2003, **327**:659-669.
44. Nagae M, Yamaguchi Y: **Three-dimensional structural aspects of protein-polysaccharide interactions.** *Int J Mol Sci* 2014, **15**:3768-3783.
45. Robertus J: **The structure and action of ricin, a cytotoxic N-glycosidase.** *Semin Cell Biol* 1991, **2**:23-30.
46. Taylor ME: **Evolution of a family of receptors containing multiple motifs resembling carbohydrate-recognition domains.** *Glycobiology* 1997, **7**:R5-R8.
47. Fujimoto Z, Kuno A, Kaneko S, Kobayashi H, Kusakabe I, Mizuno H: **Crystal structures of the sugar complexes of *Streptomyces olivaceoviridis* E-86 xylanase: sugar binding structure of the family 13 carbohydrate binding module.** *J Mol Biol* 2002, **316**:65-78.
48. Fouquaert E, Peumans WJ, Gheysen G, Van Damme EJM: **Identical homologs of the *Galanthus nivalis* agglutinin in *Zea mays* and *Fusarium verticillioides*.** *Plant Physiol Biochem* 2011, **49**:46-54.
49. Ghequire MGK, Loris R, De Mot R: **MMBL proteins: from lectin to bacteriocin.** *Biochem Soc Trans* 2012, **40**:1553-1559.
50. Ghequire MGK, Garcia-Pino A, Lebbe EKM, Spaepen S, Loris R, De Mot R: **Structural determinants for activity and specificity of the bacterial toxin LipA.** *PLoS Pathog* 2013, **9**:e1003199.

Two recent structures confirm that these bacteriocins have B-lectin folds as predicted from sequence comparison and the results demonstrate that they function in sugar binding.

See annotation to Ref [50\*].

51. McCaughey LC, Grinter R, Josts I, Roszak AW, Waløen KI, Cogdell RJ, Milner J, Evans T, Kelly S, Tucker NP *et al.*: **Lectin-like bacteriocins from *Pseudomonas* spp. utilise D-rhamnose containing lipopolysaccharide as a cellular receptor.** *PLoS Pathog* 2014, **10**:e1003898.
52. Rigden DJ, Mello LV, Galperin MY: **The PA14 domain, a conserved all- $\beta$  domain in bacterial toxins, enzymes, adhesins and signaling molecules.** *Trends Biochem Sci* 2004, **29**:335-339.
53. Schallus T, Jaeckh C, Fehér K, Palma AS, Liu Y, Simpson JC, Mackeen M, Stier G, Gibson TJ, Feizi T *et al.*: **Malectin: a novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate play in the early steps of protein N-glycosylation.** *Mol Biol Cell* 2008, **19**:3404-3414.
54. Lindner H, Müller LM, Boisson-Dernier A, Grossniklaus U: **CrRLK1L receptor-like kinases: not just another brick in the wall.** *Curr Opin Plant Biol* 2012, **15**:659-669.

The potential roles of sugar-binding domain in receptor-like protein kinases in plants are reviewed critically.



55. Powlesland AS, Ward EM, Sadhu SK, Guo Y, Taylor ME, Drickamer K: **Novel mouse homologs of human DC-SIGN: widely divergent biochemical properties of the complete set of mouse DC-SIGN-related proteins.** *J Biol Chem* 2006, **281**:20440-20449.
56. Padler-Karavani V, Hurtado-Ziola N, Chang YC, Sonnenburg JL, Ronaghy A, Yu H, Verhagen A, Nizet V, Chen X, Varki N *et al.*: **Rapid evolution of binding specificities and expression patterns of inhibitory CD33-related Siglecs in primates.** *FASEB J* 2014, **28**:1280-1293.
57. Eisen DP: **Mannose-binding lectin deficiency and respiratory tract infection.** *J Innate Immun* 2010, **2**:114-122.
58. Venkatraman Girija U, Gingras AR, Marshall JE, Panchal R, Sheikh MA, Gál P, Schwaeble WJ, Mitchell DA, Moody PC, Wallis R: **Structural basis of the C1q/C1s interaction and its central role in assembly of the C1 complex of complement activation.** *Proc Natl Acad Sci U S A* 2013, **110**:13916-13920.
59. Alter A, de Léséleuc L, Thuc NV, Thai VH, Huong NT, Ba NN, Cardoso CC, Grant AV, Abel L, Moraes MO *et al.*: **Genetic and functional analysis of common MRC1 exon 7 polymorphisms in leprosy susceptibility.** *Hum Genet* 2010, **127**:337-348.
60. Li H, Yu X-M, Wang J-X, Hong Z-H, Tang NL-S: **The VNTR polymorphism of the DC-SIGNR gene and susceptibility to HIV-1 infection: a meta-analysis.** *PLoS ONE* 2012, **9**:e42972.
61. Feinberg H, Rowntree TJW, Tan SLW, Drickamer K, Weis WI, Taylor ME: **Common polymorphisms in human langerin change specificity for glycan ligands.** *J Biol Chem* 2013, **288**:36762-36771.
- Genomic, glycan array and structural information on polymorphic forms of the receptor langerin suggest that sequence variation leading to altered sugar-binding activity may be selected for in the evolution of receptors of the innate immune system.
62. Wu Z, Lou Y, Lu L, Liu Y, Chen Q, Chen X, Jin W: **Heterogeneous effect of two selectin gene polymorphisms on coronary disease risk: a meta-analysis.** *PLOS ONE* 2014, **9**:e88152.
63. Preston RC, Rabbani S, Binder FPC, Moes S, Magnani J, Ernst B: **Implications of the E-selectin S128R mutation for drug discovery.** *Glycobiology* 2014, **24**:592-601.
- A careful re-analysis of variant forms of E-selectin suggests that changes in the epidermal growth factor domain adjacent to the carbohydrate-binding domain do not alter the sugar-binding characteristics of this receptor.