Synthetic lectins

Anthony P. Davis*

Received 19th May 2009, Accepted 7th July 2009
First published as an Advance Article on the web 5th August 2009
DOI: 10.1039/b909856a

Carbohydrate recognition presents a difficult challenge to supramolecular chemists, especially in the natural medium of water. After two decades of research, it has at last been possible to develop biomimetic receptors which perform well in aqueous solution. The "temple" family of carbohydrate receptors bind substrates with all-equatorial substitution patterns (*e.g.* β-glucosyl, β-GlcNAc, β-cellobiosyl) in a manner which is quite similar to carbohydrate-binding proteins (lectins). Affinities match some lectin–carbohydrate interactions, and selectivities are high. These "synthetic lectins" have been used to elucidate the role of water in carbohydrate recognition, and may have potential as research tools for glycobiology.

Introduction

Biology sets a hard example for molecular designers. Encouragingly, it demonstrates that molecules can possess remarkable functionality. On the other hand, its success can seem intimidating. Biological designs are so competent and sophisticated that they seem almost magical. If we could match biology with synthetic, abiotic systems we could accomplish remarkable feats of molecular engineering. We could also throw light on the operation of biological molecules. However, all experience suggests that this will be very difficult.

Chemists enjoy a challenge, and it is not surprising that they have been drawn towards biomimicry. This article highlights one goal out of many, the mimicry of biological carbohydrate recognition by lectins (carbohydrate-binding proteins). The problem has several attractions. Firstly it is difficult, among the most exacting in molecular recognition. Carbohydrates possess arrays of hydroxyl groups, and are fairly similar to clusters of water molecules. Water is the biological solvent, and thus the main competitor for a carbohydrate binding site. A receptor must be able to distinguish between water and saccharide, rejecting the former and binding the latter (Fig. 1). Moreover recognition should be selective, and the differences between carbohydrates are quite subtle. For example, a receptor should be able to discriminate between glucose **1** and galactose **2** or mannose **3**. These targets are almost exactly the same size and possess the same set of functional groups, differing only in stereochemistry at single chiral centres.

School of Chemistry, University of Bristol, Cantock's Close, Bristol, UK BS8 1TS. E-mail: Anthony.Davis@bristol.ac.uk; Fax: +44 (0) 117 929 8611; Tel: +44 (0) 117 954 6334

Antony P. Davis

Tony Davis gained a B.A. in Chemistry from Oxford University in 1977, then stayed on for a D.Phil. under Dr. G. H. Whitham and postdoctoral research with Prof. J. E. Baldwin. In 1981 he moved to the ETH Zürich to work with Prof. A. Eschenmoser, then in 1982 was appointed as a Lecturer in Organic Chemistry at Trinity College, Dublin. In September 2000 he moved to the University of Bristol, where he is Professor of Supramolecular Chemistry. His principal research interests are the biomimetic recognition of carbohydrates, anion recognition and transport, and the crystal engineering of nanoporous solids.

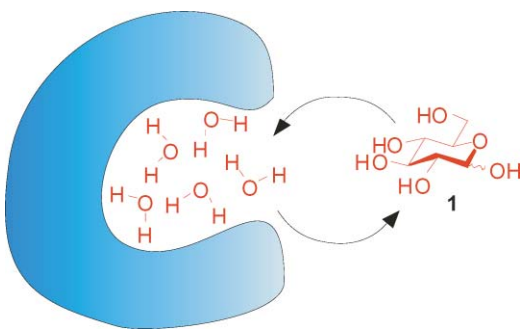
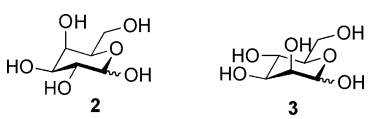


Fig. 1 The challenge of biomimetic carbohydrate recognition. The receptor must reject one set of hydroxyl groups (water molecules) in favour of another (the carbohydrate). The water molecules are present in great excess.



Secondly, carbohydrate recognition is an important natural phenomenon.[1] Saccharides play many roles in biology. They are used on a large scale as fuels (*e.g.* glucose and starch) and structural materials (*e.g.* cellulose and chitin), but they also perform more delicate tasks. They are the most information-rich of biological macromolecules, in that 6 monomers can yield $>10^{12}$ oligomeric structures (compared to 4096 for nucleotides and 6×10^7 for peptides).[2] This variability is exploited in labelling schemes for proteins and cells. The recognition of these labels mediates protein folding and trafficking,[3] cell–cell recognition,[4]

infection by pathogens,[5] tumour metastasis,[6] and many aspects of the immune response. By mimicking carbohydrate recognition, we can aspire to learn more about it; for example, if it is intrinsically so difficult (Fig. 1), how does nature succeed?†

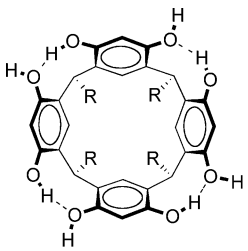
Thirdly, synthetic carbohydrate receptors (synthetic lectins) would certainly have applications if sufficiently effective. They could complement natural lectins in biological research, and could perhaps be used therapeutically (for example, by suppressing infection or inflammation).[7] There are clear possibilities in diagnostics, where a synthetic lectin for glucose would be especially interesting. In principle, this could be incorporated in an implantable blood glucose monitor capable of continuous output. Such a device would be of great benefit in the management of diabetes.

Supramolecular chemists have been publishing on carbohydrate recognition since the late 1980s. Two approaches can be distinguished. One set of researchers, following Shinkai,[8a] have exploited the reversible reaction of diols with boronic acids to give cyclic boronates.[8] This strategy is effective and probably a good route to applications, but it is not biomimetic, and is therefore outside the scope of the present article. The second approach exploits non-covalent interactions such as hydrogen bonding and CH–π interactions.[8c,9] Much of this work has avoided the key problem (Fig. 1) by using organic solvents. However, there have been attempts at true lectin mimicry, *i.e.* the binding of carbohydrates in water through non-covalent interactions with useful affinities and selectivities. Mostly these have served to illustrate the difficulty of the problem, but very recently there has been some interesting progress. This article focuses on one strand of research which has proved especially encouraging. For some substrates at least, it is now possible to make receptors which act very much like lectins. They can serve as realistic models for theoretical studies and, with further development, applications in biology may well be possible.

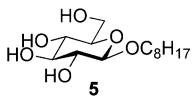
† It should be noted, however, that protein–carbohydrate interactions are often relatively weak, as illustrated later in this article.

Carbohydrate recognition in organic solvents—walk first, run later

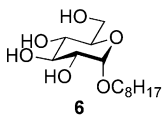
Even without competition from water, carbohydrates present quite challenging targets. They are larger than many traditional substrates of supramolecular chemistry (*e.g.* inorganic cations and anions) and considerably more complex. It is not surprising that early studies sidestepped the "water issue" and employed less competitive media. In non-polar solvents, hydrogen bonding can be very effective, especially when multiple interactions are possible. Architectures which surround or span a saccharide with H-bond donors and/or acceptors are likely to be successful. This was first demonstrated in 1988 by Aoyama's calixarene **4**,[10] which extracted certain carbohydrates from water into CCl_4[10a] and was shown to bind octyl glucosides **5** and **6** in $CDCl_3$.[10b]



4 R = $C_{11}H_{23}$



5



6

There followed a series of papers, from many groups and extending to the present day, reporting carbohydrate binding by polar interactions in organic media.[11] Fig. 2 shows a very limited selection of the systems used, encompassing the authors' "cholaphane" **7**,[11a,b] Anslyn's polyaza cleft **8**,[11e] Diederich's and Hamilton's phosph(on)ates **9**[11g,o] and **10**,[11i] Mazik's tris-pyrimidine

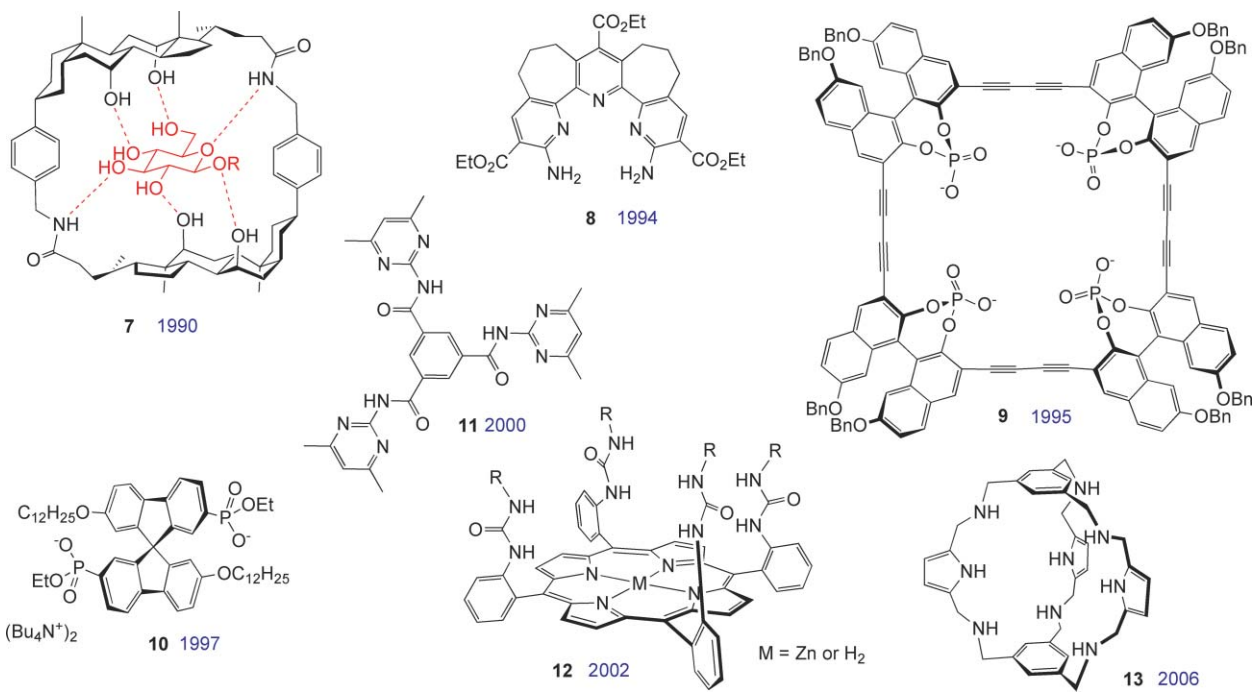


Fig. 2 Synthetic carbohydrate receptors employed in organic solvents, with dates of publication. All are presumed to bind *via* multiple hydrogen bonds, as shown schematically for **7**.

11,[11m] Bonar-Law's[11q] and Kim's[11r] ureidoporphyrins **12** and Roelens' tris-pyrrolic cage **13**.[11x] Some of these receptors were capable of high affinities. For example, one variant of **12** bound **5** with $K_a = 2 \times 10^7$ M^{-1} in $CHCl_3$,[11r] while tetraphosphate **9** gave $K_a = 5200$ M^{-1} with the same substrate in a much more competitive medium (CD_3CN/CD_3OD, 98:2).[11o] Selectivities were also promising. Many systems showed useful diasteroselectivities. Receptor **13**, for example, bound **5** with $K_a = 5 \times 10^4$ M^{-1} in $CDCl_3$ while showing no measurable affinity for α-anomer **6**.[11x] Enantioselectivity was also observed for some chiral receptors. Thus, **7** and **10** bound **5** and **ent-5** in ratios of 3:1 and 5:1 respectively.[11b,i]

This work has been useful and interesting in exploring some aspects of carbohydrate recognition. However, biological relevance is reduced by the use of unnatural solvents, especially as the role of the solvent is one of the most important specific questions (see later). The potential for applications is also limited. The carbohydrates of interest occur in water and, in most cases, will not even dissolve in many organic solvents. Indeed the substrates for use with organic-soluble receptors must be carefully chosen. Lipophilic glycosides such as **5** and **6** have proved suitable, but not many such compounds are commercially available. Substrate synthesis can be a necessary but unwelcome task for researchers in this area.

It is clearly desirable, therefore, to translate success in organic solvents into systems which operate in water. In the mid 1990s the author's group resolved to make the attempt. The work on cholaphanes such as **7** was not, however, especially encouraging. The steroidal framework of **7** is notable for its lipophilic exterior, and solubilising the system in water would not be straightforward. Also discouraging were extraction studies in aqueous–organic 2-phase systems. If a receptor could extract carbohydrates from water, one might infer that its binding site could displace a shell of solvating water molecules. Sadly the cholaphanes failed to extract glucose, and only succeeded with methyl glucoside (which is more lipophilic) at very high concentrations.[12] The cholaphane affinities peaked at ~3000 M^{-1} in $CDCl_3$ and, apparently, this was not enough to compete with water. We therefore targeted a new system which might show greater affinities in all media, and which should be more compatible with operation in water.

"Temple" carbohydrate receptors; early work in organic solvents

For the design of our new receptor architecture, we planned to take the most rational approach possible. Experience, theory and the observation of natural systems suggest that the key to recognition is complementarity. A receptor should be the correct size and shape for its target, should provide appropriate matches for polar functionality (*e.g.* H-bond donor for H-acceptor, *etc.*), and should also match apolar surfaces in host and guest. This last requirement is perhaps less important in apolar media (although it can certainly do no harm). However, it is clearly necessary in water, where hydrophobic interactions are likely to make important contributions.

The process of receptor design should therefore involve (a) choosing a substrate, (b) constructing (on paper, or *in silico*) a complementary binding site, with the substrate as a virtual template, (c) ensuring that the proposed structure retains its shape sufficiently in the absence of the substrate, and (d) adjusting properties, notably solubility, by tuning of peripheral groups. This process, it should be said, had not been undertaken for cholaphane **7**, where the exploitation of a novel starting material (the steroid cholic acid) had been a major driving force.[13] Regarding the substrate, glucose was the obvious choice. It is the most common carbohydrate, and indeed the most common organic molecule on earth when polymeric forms are included. It is also the target for a major practical application, blood analysis for diabetics.

Glucose possessed a further attraction in that its structure, though fully asymmetric, has a certain regularity. As the β anomer it is the "all-equatorial" carbohydrate, with polar groups directed away from the centre and two small patches of hydrophobic, axial C-H groups. To a very rough approximation, it may be represented as a squat cylinder with hydrophobic ends and a polar circumference (Fig. 3a). Complementarity to this simplified object might not be so hard to achieve. Two apolar surfaces set parallel to each other, separated by rigid polar spacers, would create a cavity which might bind glucose well in both organic and aqueous media (Fig. 3b). The apolar surfaces should be aromatic to take advantage of the CH–π interaction.[14] In cartoon form the architecture is reminiscent of a classical temple, providing us with a name for this family of structures. Biological systems offered encouragement. In particular, the crystal structure of an *E. coli* carbohydrate sensing and transport protein showed a similar arrangement (Fig. 4).[15] Glucose is sandwiched between aromatic residues, while being held by a comprehensive network of hydrogen bonds.

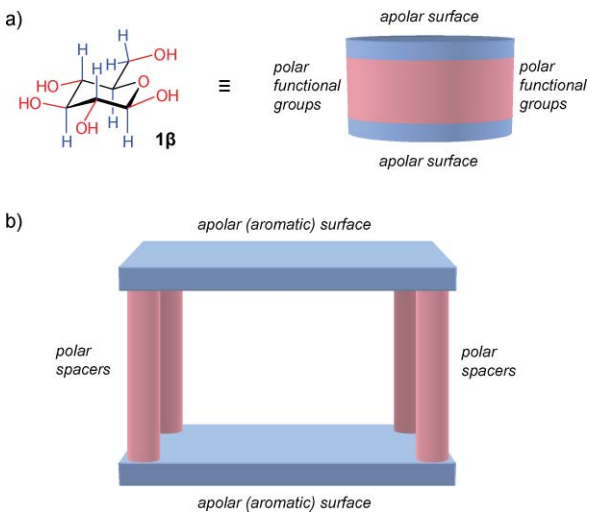


Fig. 3 (a) Analysis of the β-glucose structure **1β** as regions of contrasting polarity. (b) The complementary "Temple" architecture for a glucose receptor.

Our first design for a carbohydrate-binding "temple" was macrotricycle **15**.[16] The apolar roof and floor were provided by biphenyl units, while the polar pillars were to be isophthalamides. With 8 secondary amides in the pillars, hydrogen bonding to the substrate should be more effective than in **7**.[17] Solubility in organic media should be promoted by the external pentyl esters, and hydrolysis of these esters should give a tetracarboxylate for future study in water. Modelling of the methyl ester analogue **16** confirmed that a β-glucosyl unit should fit neatly inside, making up to 6 intermolecular hydrogen bonds and several CH–π interactions

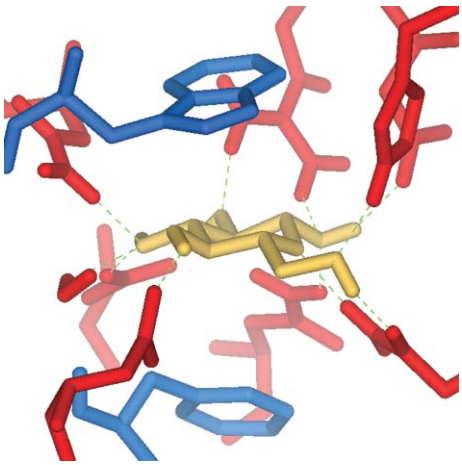


Fig. 4 Binding site of the *E. coli* galactose chemoreceptor protein, with glucose substrate (yellow).[15] Aromatic tryptophan and phenylalanine residues are shown in blue, polar residues in red.

(Fig. 5). Importantly, modelling of the empty receptor showed that collapse to a closed conformation was disfavoured. Although the spacer units could fold inwards to occupy the binding site, this was only possible if two or three amides adopted high-energy *syn* conformations.

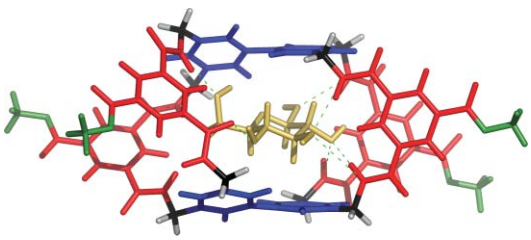
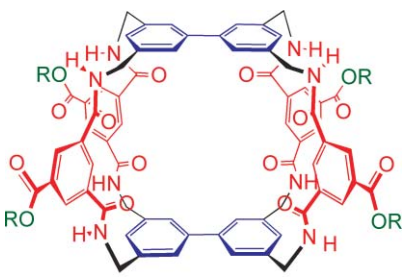
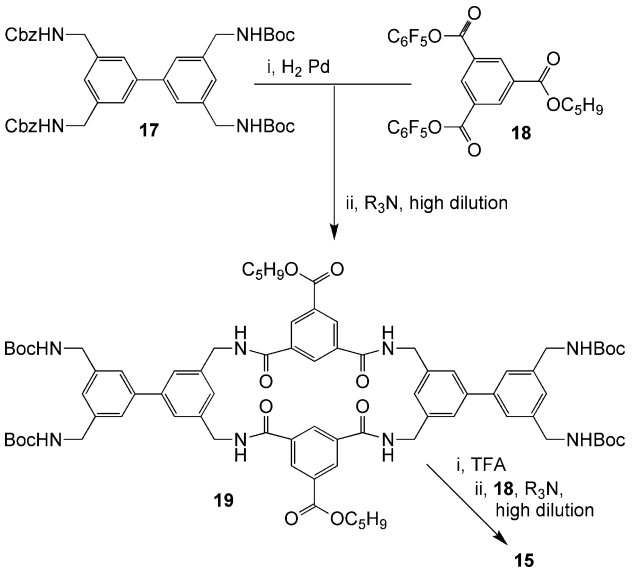


Fig. 5 Baseline conformation from a Monte Carlo Molecular Mechanics study of **16** + β-glucose **1**β. Atoms of **16** are coloured to match the formula. The glucose molecule is shown as yellow.



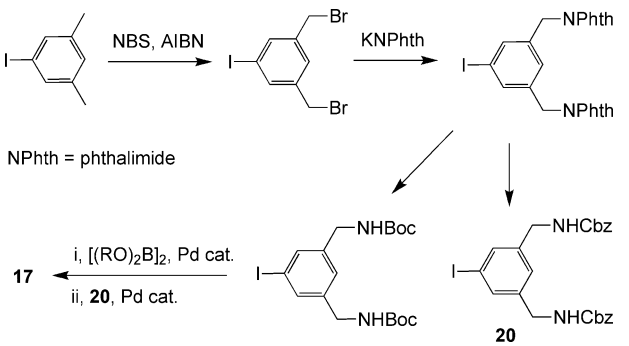
15 R = C_5H_9
16 R = CH_3

Synthetic accessibility is, of course, another major design criterion. Receptor **15** was successfully constructed from two components, biphenyl **17** and triester **18**, as shown in Scheme 1.[18] Yields were improved by two factors. Firstly, the 1+1 cyclisation of **17** with **18** cannot take place because the product is too strained. The 2+2 cyclisation product **19** is the smallest possible, and is therefore favoured at high dilution. Secondly, the cyclisation of **19** is simplified by free rotation about the biphenyl Ar–Ar bonds. If "roof" and "floor" were composed of rigid units, the macrocycle could react with **18** to give two regioisomers, only one of which



Scheme 1 The assembly of **15** from **17** and **18**.

could proceed to a tricyclic product. The biphenyl bond rotation removes this problem—**19** + **18** can give just one cyclised product. Biphenyl **17** was prepared initially *via* a sequence involving Stille coupling,[18] although more recently Suzuki–Miyaura methodology has been preferred (Scheme 2).[19]



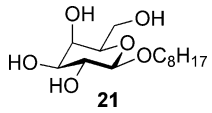
Scheme 2 Preparation of biphenyl **17**.

The binding properties of **15** proved encouraging. Measurements were performed by NMR in $CDCl_3/CD_3OD$ (92:8), and also by fluorescence spectroscopy in $CHCl_3$. The results are summarised in Table 1. In chloroform, the temple receptor bound β-glucoside **5** ~100 times more strongly than had cholaphane **7**. Affinities in $CDCl_3/CD_3OD$ (92:8) were lower, but still quite impressive considering the competitive nature of this solvent system. Selectivity for **5** *vs.* galactoside **21** was quite good (~4:1), and the preference for β *vs.* α glucoside was excellent (~40:1).

Table 1 Binding constants (M^{-1}) measured to receptor **15**

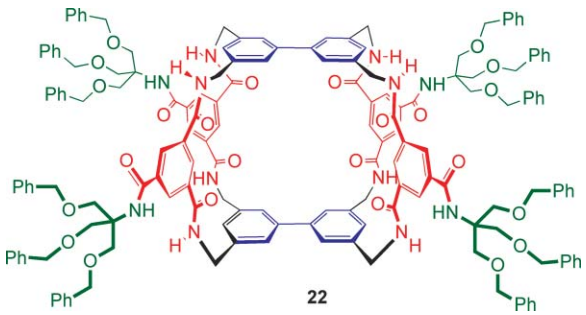
Substrate	$CDCl_3$–CD_3OH (92:8)[a]	$CHCl_3$[b]
β-D-glucoside **5**	980	300,000
α-D-glucoside **6**	20	13,000
β-D-galactoside **21**	220	110,000

[a] Determined by ^1H NMR titration at 303 K. [b] Determined by fluorescence titration.



21

Temple **15** was also capable of interacting with solid glucose. When glucose was stirred with the receptor in chloroform, ~0.9 equivalents of the carbohydrate were solubilised. Phase transfer experiments in chloroform/water were less successful, as the receptor seemed to precipitate at the interface. However, we later prepared the receptor **22** with more substantial solubilising groups (see also next section), and this was more effective.[19] For example, from a 0.5 M aqueous solution of glucose, **22** was able to extract ~0.5 equivalents into chloroform. Both galactose and mannose were extracted in far smaller amounts.

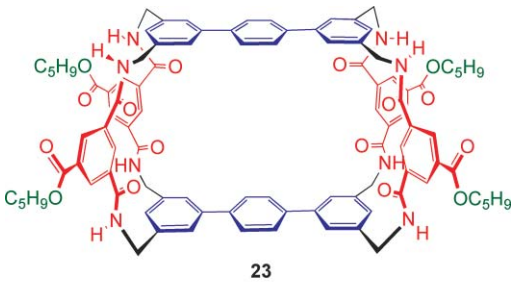


22

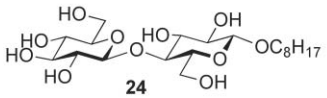
Carbohydrates exert their biological effects mostly in combination with each other, as part of complex oligosaccharide units. Oligosaccharides are therefore very relevant targets for biomimetic recognition. We realised that the synthesis of **15** could be adapted to yield **23**, and that this "extended temple" should be complementary to all-equatorial disaccharides (*i.e.* cellobiosides such as **24**). Receptor **23** was prepared, and was found to bind **24** with $K_a = 7000\ M^{-1}$ in $CDCl_3/CD_3OD$ (92:8).[20] Extending receptor and target had thus raised affinity by a modest amount. More important, perhaps, was the selectivity of **23**. The extended temple was tested against glucosides **5** and **6**, lactoside **25** and maltosides **26** and **27**. None showed evidence of complex formation, either by ^{1}H NMR or by fluorescence spectroscopy. The contrast between cellobioside **24** and lactoside **25** is remarkable; inversion of just one stereocentre out of 10 was sufficient for complete disruption of binding.

Temples in water; the first "synthetic lectins"?

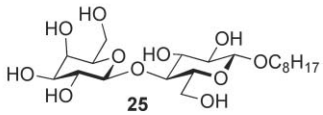
The temple receptors had proved quite successful in organic media. The next step was to test the concept in water. The literature warned against high expectations. Although a number of systems have been reported to bind carbohydrates in water,[21] affinities have tended to be very low. This is especially true for cases where complex formation is well-characterised.[22] A representative example is the calixarene **28**, a water-soluble relative of **4**.[21a] This molecule was found to bind fucose **29** and deoxyribose **30** with K_a = 6 and 5 M^{-1} respectively. These substrates are arguably less challenging than many other carbohydrates, as they are relatively hydrophobic. Binding to the common hexoses **1–3** was not reported.



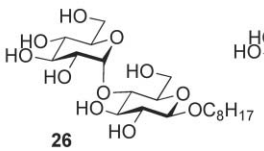
23



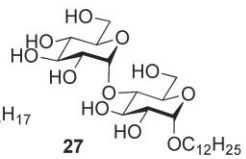
24



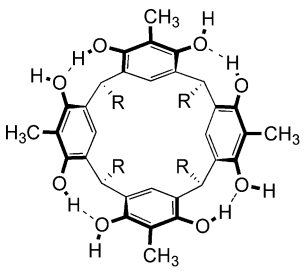
25



26

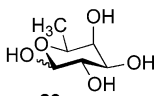


27

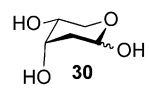


28

R = $CH_2CH_2SO_3Na$

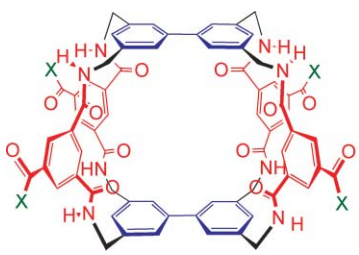


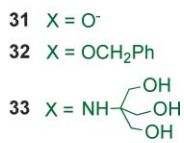
29



30

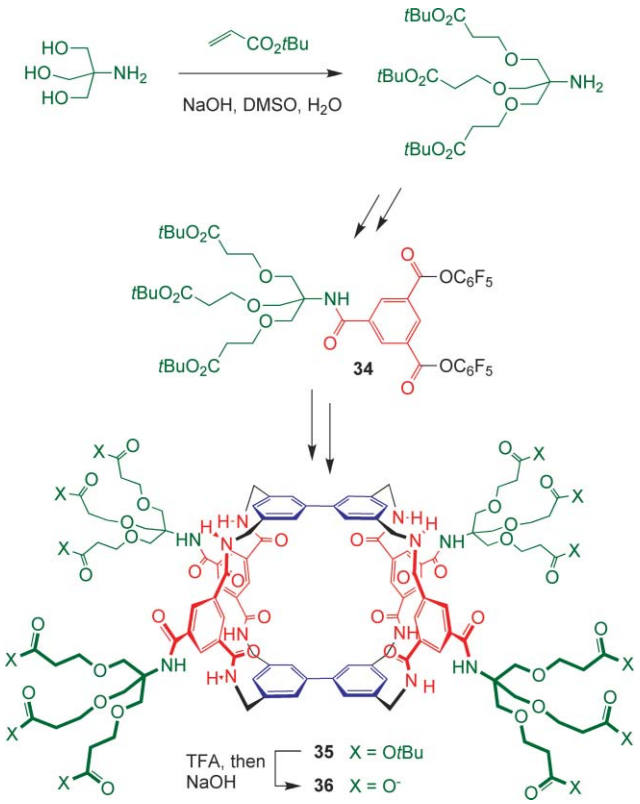
To study the temple architecture in water we needed a water-soluble variant. Moreover we wished to use NMR which, through effects such as intermolecular shielding and NOE, gives unambiguous information on binding. We therefore needed quite high solubility (~1 mM) and a system which did not show significant aggregation at these concentrations. It proved quite difficult to fulfil these criteria. Tetra-anion **31** was prepared *via* hydrogenolysis of **32**, but gave broadened ^{1}H NMR spectra in water. Dodecaol **33** was obtained by debenzylation of **22**, but proved similarly unsuitable.[23]





31 X = O^-
32 X = OCH_2Ph
33 X = NH-

Finally the spacer unit **34**, with 3 masked carboxylate groups, was prepared as shown in Scheme 3 and incorporated



Scheme 3 The synthesis of water-soluble temple **36**.

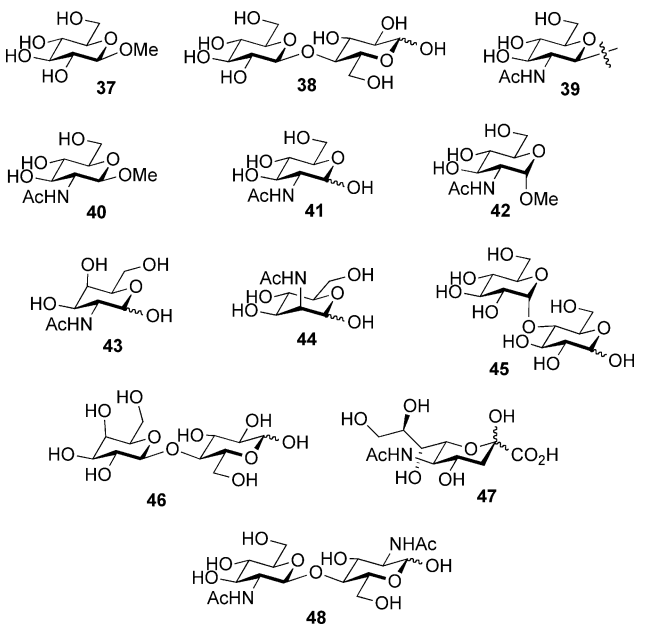
in **35**. Deprotection with TFA and dissolution in aqueous NaOH gave dodecacarboxylate **36**. Receptor **36** was freely soluble in water and, thankfully, gave well-resolved ^{1}H NMR spectra.[24]

As hoped, the binding of carbohydrates to **36** could be detected, and quantified, by ^{1}H NMR titrations. A first set of studies pointed to glucose **1**, methyl β-D-glucoside **37** and cellobiose **38** as good substrates, in accordance with the designed preference for all-equatorial substitution patterns.[24] Binding constants were low (*e.g.* 9 M^{-1} for **1**), but significant considering the challenge of binding hexoses in water. However, a second series of experiments altered the picture. These were inspired by the realisation that β-*N*-acetylglucosaminyl (β-GlcNAc, **39**), a common (all-equatorial) natural saccharide unit, was also a potential substrate. It transpired that temple **36** is a very good and remarkably selective receptor for β-GlcNAc.[25] Table 2 lists the binding constants measured for **36** to a wide range of carbohydrate derivatives. The methyl glycoside **40**, the simplest anomerically fixed representative of β-GlcNAc, was bound with $K_a = 630$ M^{-1} (confirmed by isothermal titration calorimetry). *N*-Acetylglucosamine itself was complexed more weakly, but NMR studies (see below) indicated that only the β-anomer was bound. Taking account of the predominance of the α anomer (α:β = 64:36), K_a for the β form could be estimated at ~150 M^{-1}. In comparison, the β-glucosides are quite poor substrates. For example, the selectivity for GlcNAcβ-OMe **40** *vs.* methyl β-D-glucoside **37** was 23:1. The other carbohydrates tested were hardly bound at all.

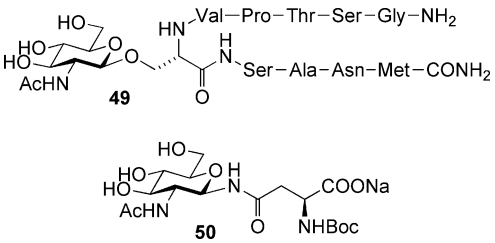
Table 2 Binding constants (K_a, M^{-1}) measured to receptor **36** in water, with values to the lectin Wheat Germ Agglutinin (WGA) for comparison

Substrate	**36**[a]	WGA[b]
D-GlcNAcβ-OMe **40**	630	730
D-GlcNAc **41** (α:β = 64:36)	56	410
methyl β-D-glucoside **37**	28	
GlcNAcα-OMe **42**	24[c]	480
D-cellobiose **38**	17	
D-glucose **1**	9	
2-deoxy-D-glucose	7	
methyl α-D-glucoside	7	
D-xylose	5	
D-ribose	3	
D-galactose **2**	2	
L-fucose **29**	2	
N-acetyl-D-galactosamine **43**	2	60
N-acetyl-D-mannosamine **44**	2	60
D-arabinose	2	
D-lyxose	≤2	
D-mannose **3**	≤2	
L-rhamnose	≤2	
D-maltose **45**	≤2	
D-lactose **46**	≤2	
N-acetyl-D-muramic acid	0[d]	
N-acetyl-D-neuraminic acid **47**	0[d]	560
N,*N*′-diacetyl-D-chitobiose **48**	0[d]	5300

[a] Determined by ^{1}H NMR titration unless otherwise stated. [b] See ref. 25. [c] Measured by induced circular dichroism (ICD). [d] No change in spectrum on addition of carbohydrate.



By good fortune, it turns out that β-GlcNAc is a uniquely interesting target. Placed on the serine and threonine hydroxyls of proteins, it is an important regulatory post-translational modification (the O-GlcNAc modification).[26] To test whether **36** could bind β-GlcNAc in this environment, the glycopeptide **49** was prepared and employed as a substrate. The complex **36**·**49** did indeed form, with $K_a = 1040$ M^{-1}.[25] The N-linked derivative was bound very weakly ($K_a = 4$ M^{-1}), and the β-GlcNAc terminated disaccharide *N*,*N*′-diacetylchitobiose **48** was not complexed at all, so the selectivity of **36** seems to be very tightly defined.



Lectins are widely used in glycobiological research. For specific binding of GlcNAc (including the O-GlcNAc modification) the standard tool has been Wheat Germ Agglutinin (WGA). It is therefore interesting to compare **36** and WGA. As shown in Table 2, **36** roughly matches the affinity of WGA for simple β-GlcNAc derivative **40** and is far more selective for this type of unit. WGA has significant affinity for α-GlcNAc derivative **42**, and the other *N*-acetylamino sugars **43**, **44** and **47**. It binds *N*,*N*′-diacetylchitobiose **48** very strongly indeed. Receptor **36** is thus complementary to WGA, and may be more suitable for specific purposes such as binding O-GlcNAc-modified peptides. Given its overall performance, it seems quite justifiable to describe **36** as a "synthetic lectin".[27]

Finally, NMR studies provided unusual levels of structural information on binding by **36**, helped by the fact that some complexes showed slow dissociation on the ^{1}H NMR timescale. One such case was **36**·**41**, where the spectrum showed binding to just the β anomer. Another was **36**·**40**, for which NOE data allowed an unambiguous structural determination (Fig. 6). The carbohydrate was found to lie between the biphenyl units, with axial CH groups making CH–π contacts as expected. The NHAc group was positioned between two spacer groups in a narrow portal of the cavity. This detailed structure raises the possibility of rational tuning of the design to improve affinity and selectivity.

As discussed earlier, the temple design can be extended to target all-equatorial disaccharide units, as in cellobiose **38** and *N*,*N*′-diacetylchitobiose **48**. Receptor **23** had been successful in organic solvents, but it was feared a water-soluble analogue might collapse in a twisting motion, driven by hydrophobic interactions. Instead the receptor **53** was designed, based on meta-terphenyl units and employing a fifth isophthalamide spacer.[28] Modelling predicted

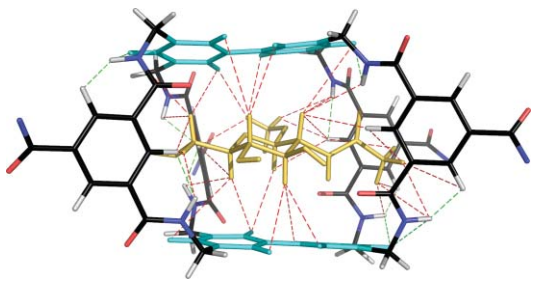
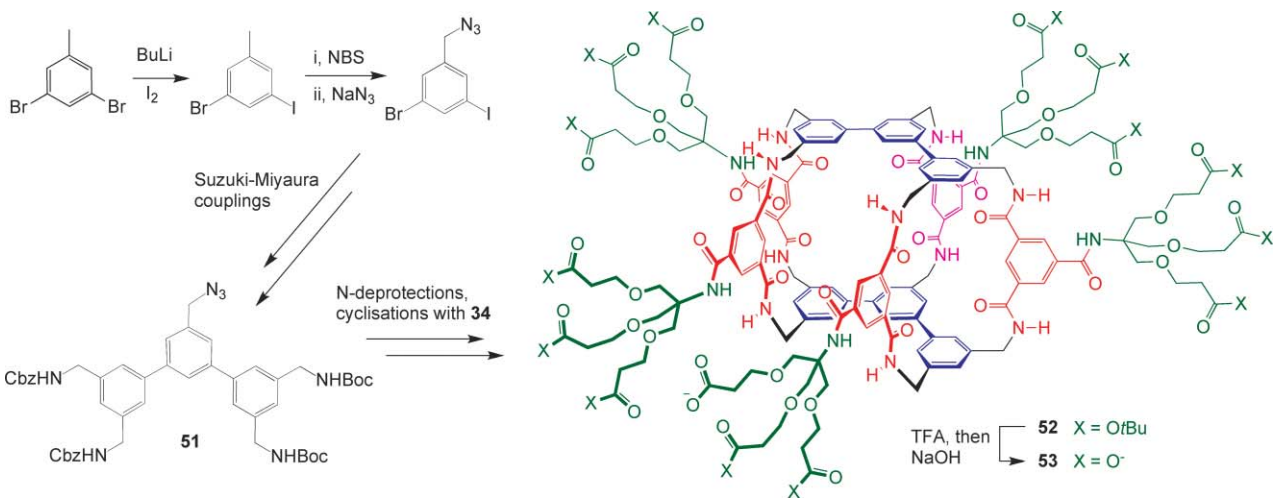


Fig. 6 NMR structure for the complex between **36** and GlcNAcβ-OMe **40**. The biphenyl units are highlighted in cyan, and carbohydrate **40** is shown as yellow. Intramolecular and intermolecular NOE contacts are shown as green and red broken lines respectively. The water-solubilising tricarboxylate groups are omitted.

that the terphenyl surfaces of **53** would be held apart in all accessible conformations. **53** was prepared as indicated in Scheme 4, *via* differentially-protected intermediate **51**. Like **36** it dissolved freely in water to give well-resolved NMR spectra.

Binding to carbohydrates could be studied by ^{1}H NMR, induced circular dichroism (ICD) and fluorescence titrations, giving the results shown in Table 3. At least two methods were used for each substrate, and agreement was generally good, so the K_a values should be especially reliable. Once again the system showed good affinities and excellent selectivities for the intended (all-equatorial) substrates. Cellobiose **38** and methyl β-D-cellobioside **54** were bound with $K_a \approx 600$ and 900 M^{-1} respectively. Xylobiose **55** ($K_a \approx 260$ M^{-1}) and *N*,*N*′-diacetyl-D-chitobiose **48** ($K_a \approx 120$ M^{-1}) were complexed somewhat less strongly. The non-targeted disaccharides **45**, **46** and **56–59** showed low affinities, a factor of ~50 less than **38**. Like its organic-soluble relative **23**, receptor **53** distinguished readily between cellobiosyl and lactosyl despite the fact that these units differ at just one asymmetric centre. Monosaccharides were also bound weakly, presumably because they are too small to contact all parts of the cavity. The selectivity for cellobiosyl was confirmed in a competition experiment, in which **38** (9 mM) was added to **53** in the presence of 8 other carbohydrates, each at 20 mM. Complex **52**·**38** was observed to form almost quantitatively.



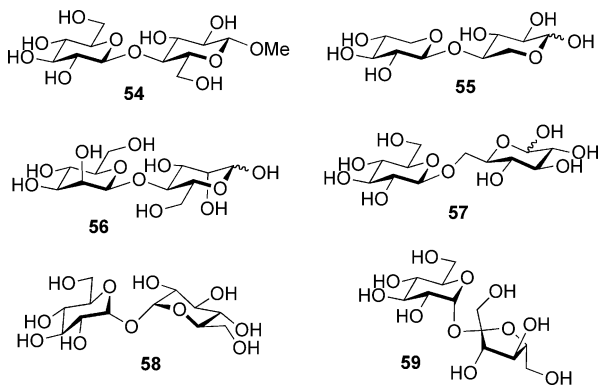
Scheme 4 Synthesis of "extended temple" receptor **53**.

Table 3 Binding constants (K_a, M^{-1}) measured to "extended temple" **52** in water, as measured by ^{1}H NMR, ICD and fluorescence titrations[28]

Substrate	^{1}H NMR	ICD	Fluorescence
D-cellobiose **38**	600	580	560
methyl β-D-cellobioside **54**		910	850
D-xylobiose **55**		250	270
N,*N*′-diacetyl-D-chitobiose **48**	120		120
D-lactose **46**		11	14
D-mannobiose **56**		13	9
D-maltose **45**		15	11
D-gentiobiose **57**		12	5
D-trehalose **58**	0[a]	0[a]	
D-sucrose **59**		0[a]	0[a]
D-glucose **1**	11	12	0[a]
D-ribose		0[a]	0[a]
D-GlcNAc **41**	24		19

[a] No change in spectrum on addition of carbohydrate.

The cellobiose complex **53**·**38** was also studied by isothermal titration calorimetry (ITC). As well as providing a fourth value for the binding constant (650 M^{-1}), this technique revealed the thermodynamic driving force for binding. ΔH was measured as –3.22 kcal mol^{-1} and TΔS as 0.62 kcal mol^{-1}, so complex formation was mainly enthalpy-driven with entropy playing a minor role (as often found for lectins[29]). The complex was also amenable to NMR structural investigations, as it showed slow dissociation on the ^{1}H NMR timescale. In this case only a few unambiguous NOE contacts could be determined but, when combined with Monte Carlo Molecular Mechanics (MCMM) calculations, a credible structural model could be developed (Fig. 7). As intended, the cellobiose is sandwiched between the terphenyl surfaces making a series of CH–π contacts. If lactose **46** is placed in the same position, the axial OH group pushes against the aromatic surface and severely disrupts the structure.



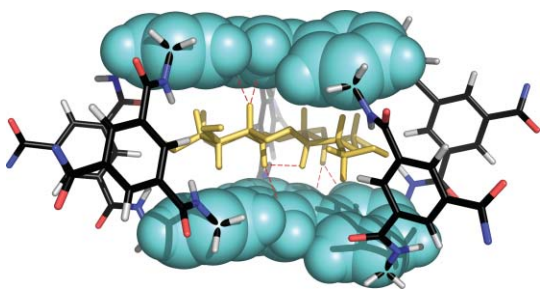


Fig. 7 Proposed structure for the complex between **52** and cellobiose **38** (β-anomer). The biphenyl units are highlighted in cyan (space-filling mode), and the cellobiose is shown as yellow. Intermolecular NOE contacts are shown as red broken lines. The water-solubilising tricarboxylate groups are omitted.

Temple receptors **36** and **53** show quite high levels of biomimicry. Both are capable of binding carbohydrates with K_a ~1000 m^{-1}, and while this is weak by general biological standards it is not unusual for lectin carbohydrate interactions (as illustrated previously for WGA).[29] They are both highly selective, and seem to operate in a manner very similar to saccharide binding proteins. They may therefore be seen as realistic lectin models, with potential to throw light on issues concerning natural carbohydrate recognition. One such issue is the role of solvent. As discussed earlier, the presence of water renders carbohydrate binding intrinsically difficult (Fig. 1). The question arises; given that binding does take place, what is the driving force? According to one view, the interaction is essentially polar in nature, driven by exceptionally favourable hydrogen bonding between protein and carbohydrate.[30] On the other hand, it has been suggested that carbohydrate-binding sites may not be well-hydrated due to their amphiphilic nature, despite containing many polar groups. If so, binding could be largely hydrophobic, driven by the displacement of high-energy water molecules.[31]

This problem is most easily addressed by varying the solvent. Solvophobic interactions are much stronger in water than other solvents so, if they are important, moving from water to a less polar solvent should *lower* affinities. On the other hand, if binding is exclusively polar in nature, water should be the most competitive of all media. Moving to any other solvent system should therefore *raise* affinities.

The difficulty with this approach is that natural carbohydrate receptors are proteins and tend to denature in non-aqueous media. However, the temple receptors possess robust cores and can be tuned for solubility in almost any medium. Thus, **36** and **53** are soluble in water and aqueous–organic mixtures, while analogues such as **15** and **52** are compatible with organic solvents including chloroform. Using these systems, with appropriate substrates, it is therefore possible to examine the full range of solvent polarities from chloroform to water, employing a third solvent such as methanol to mediate the transition.[32] The results of one such set of experiments, employing **52** and **53**, are shown in Fig. 8. The right hand graph (Fig. 8b) brings no surprises. In

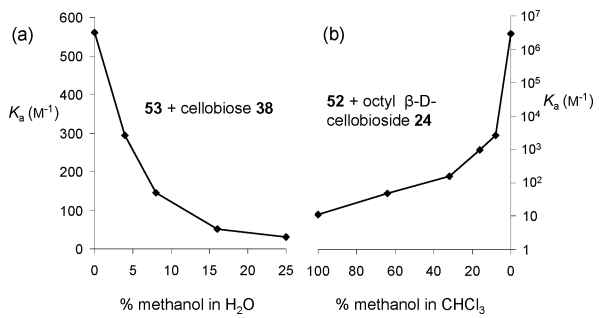


Fig. 8 Binding constants of disaccharide receptors **52** and **53** to cellobiosyl units in a series of water–methanol and methanol–chloroform solvent mixtures. (a) **53** + cellobiose **38** in water–methanol (K_a on linear scale). (b) **52** + octyl cellobioside **24** in choloroform–methanol (K_a on logarithmic scale).

the non-polar medium polar interactions dominate. The addition of methanol competes for H-bonding sites and therefore depresses affinities. However, the left hand graph (Fig. 8a) is instructive. In the aqueous medium methanol again lowers binding constants. Methanol could not depress polar interactions in water, so it must be interfering with solvophobic forces. The clear implication is that hydrophobic effects contribute substantially to carbohydrate binding in aqueous solution.

Conclusions

Protein mimicry is undoubtedly a difficult task. Operating over millions of years, biological evolution is a very efficient design mechanism and a fearsome competitor. Nonetheless, the quest should not be seen as hopeless. The chemist has greater structural scope than biology, not being limited to polypeptides derived from standard amino acids. The synthesis of complex, non-polymeric, abiotic designs is always likely to be non-trivial, but if the prediction of properties is certain enough it may be worthwhile. In this context, the work described in the present article might be thought encouraging. By following a rational design strategy, it has proved possible to develop molecules which at least bear comparison with their biological counterparts.[27,33] Fortune has surely played a role, and the freedom to shift targets (from glucose to GlcNAc) has also been important. Moreover, it is unclear at present how other families of saccharides might be addressed. However, computational methods will improve, and molecular design should become a less uncertain business. Persistence will be required but the mimicry of lectins, and by extension other proteins, no longer seems quite so unrealistic.

Acknowledgements

The author is grateful to the co-workers who have worked over more than two decades to develop synthetic carbohydrate receptors, to collaborators who have made important contributions, and to the institutions who have provided generous financial support for the research (Enterprise Ireland, the European Union, the University of Bristol and EPSRC).

References

1 Leading references: T. K. Lindhorst, *Essentials of carbohydrate chemistry and biochemistry*, Wiley-VCH, Weinheim, 2007; H. J. Gabius, H. C. Siebert, S. Andre, J. Jiménez-Barbero and H. Rudiger, *ChemBioChem*, 2004, **5**, 740; R. A. Dwek and T. D. Butters, *Chem. Rev.*, 2002, **102**, 283; C. R. Bertozzi and L. L. Kiessling, *Science*, 2001, **291**, 2357.
2 R. A. Laine, *Glycobiology*, 1994, **4**, 759; B. G. Davis, *Chem. Rev.*, 2002, **102**, 579.
3 J. Roth, *Chem. Rev.*, 2002, **102**, 285.
4 S.-i. Hakomori and K. Handa, in *Carbohydrates in chemistry and biology*, ed. B. Ernst, G. W. Hart and P. Sinay, Wiley-VCH, Weinheim, 2000, p. II 771.
5 T. Angata and A. Varki, *Chem. Rev.*, 2002, **102**, 439.
6 R. Kannagi, in *Carbohydrate-based drug discovery*, ed. C.-H. Wong, Wiley-VCH, Weinheim, 2003, p. 803.
7 J. Balzarini, *Nature Reviews Microbiology*, 2007, **5**, 583.
8 (*a*) S. Shinkai, K. Tsukagoshi, Y. Ishikawa and T. Kunitake, *J. Chem. Soc., Chem. Commun.*, 1991, 1039; (*b*) T. D. James, K. R. A. S. Sandanayake and S. Shinkai, *Angew. Chem., Int. Ed. Engl.*, 1996, **35**, 1910; (*c*) A. P. Davis and T. D. James, in *Functional Synthetic Receptors*, ed. T. Schrader, and A. D. Hamilton, Wiley-VCH, Weinheim, 2005, p. 45; (*d*) T. D. James, M. D. Phillips, and S. Shinkai, *Boronic Acids in Saccharide Recognition*, RSC, Cambridge, 2006.
9 A. P. Davis and R. S. Wareham, *Angew. Chem., Int. Ed.*, 1999, **38**, 2978; A. Lutzen, in *Highlights in bioorganic chemistry: methods and applications*, ed. C. Schmuck and H. Wennemers, Wiley-VCH, Weinheim, 2004, p. 109; M. Mazik, *Chem. Soc. Rev.*, 2009, **38**, 935.
10 (*a*) Y. Aoyama, Y. Tanaka, H. Toi and H. Ogoshi, *J. Am. Chem. Soc.*, 1988, **110**, 634; (*b*) Y. Kikuchi, Y. Tanaka, S. Sutaro, K. Kobayashi, H. Toi and Y. Aoyama, *J. Am. Chem. Soc.*, 1992, **114**, 10302.
11 Selected examples: (*a*) R. P. Bonar-Law, A. P. Davis and B. A. Murray, *Angew. Chem., Int. Ed. Engl.*, 1990, **29**, 1407; (*b*) K. M. Bhattarai, R. P. Bonar-Law, A. P. Davis and B. A. Murray, *J. Chem. Soc., Chem. Commun.*, 1992, 752; (*c*) P. B. Savage, S. K. Holmgren, J. M. Desper and S. H. Gellman, *Pure Appl. Chem.*, 1993, **65**, 461; (*d*) R. Liu and W. C. Still, *Tetrahedron Lett.*, 1993, **34**, 2573; (*e*) C.-Y. Huang, L. A. Cabell and E. V. Anslyn, *J. Am. Chem. Soc.*, 1994, **116**, 2778; (*f*) R. P. Bonar-Law and J. K. M. Sanders, *J. Am. Chem. Soc.*, 1995, **117**, 259; (*g*) S. Anderson, U. Neidlein, V. Gramlich and F. Diederich, *Angew. Chem., Int. Ed. Engl.*, 1995, **34**, 1596; (*h*) T. Mizutani, T. Kurahashi, T. Murakami, N. Matsumi and H. Ogoshi, *J. Am. Chem. Soc.*, 1997, **119**, 8991; (*i*) G. Das and A. D. Hamilton, *Tetrahedron Lett.*, 1997, **38**, 3675; (*j*) A. Bahr, A. S. Droz, M. Puntener, U. Neidlein, S. Anderson, P. Seiler and F. Diederich, *Helv. Chim. Acta*, 1998, **81**, 1931; (*k*) T. Schrader, *J. Am. Chem. Soc.*, 1998, **120**, 11816; (*l*) M. Inouye, J. Chiba and H. Nakazumi, *J. Org. Chem.*, 1999, **64**, 8170; (*m*) M. Mazik, H. Bandmann and W. Sicking, *Angew. Chem., Int. Ed.*, 2000, **39**, 551; (*n*) J. Bitta and S. Kubik, *Org. Lett.*, 2001, **3**, 2637; (*o*) A. S. Droz, U. Neidlein, S. Anderson, P. Seiler and F. Diederich, *Helv. Chim. Acta*, 2001, **84**, 2243; (*p*) S. Tamaru, S. Shinkai, A. B. Khasanov and T. W. Bell, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 4972; (*q*) K. Ladomenou and R. P. Bonar-Law, *Chem. Commun.*, 2002, 2108; (*r*) Y. H. Kim and J. I. Hong, *Angew. Chem., Int. Ed.*, 2002, **41**, 2947; (*s*) M. Segura, B. Bricoli, A. Casnati, E. M. Munoz, F. Sansone, R. Ungaro and C. Vicent, *J. Org. Chem.*, 2003, **68**, 6296; (*t*) T. Ishi-i, M. A. Mateos-Timoneda, P. Timmerman, M. Crego-Calama, D. N. Reinhoudt and S. Shinkai, *Angew. Chem., Int. Ed.*, 2003, **42**, 2300; (*u*) J. M. Fang, S. Selvi, J. H. Liao, Z. Slanina, C. T. Chen and P. T. Chou, *J. Am. Chem. Soc.*, 2004, **126**, 3559; (*v*) J. L. Hou, X. B. Shao, G. J. Chen, Y. X. Zhou, X. K. Jiang and Z. T. Li, *J. Am. Chem. Soc.*, 2004, **126**, 12386; (*w*) J. D. Lee, N. T. Greene, G. T. Rushton, K. D. Shimizu and J. I. Hong, *Org. Lett.*, 2005, **7**, 963; (*x*) O. Francesconi, A. Ienco, G. Moneti, C. Nativi and S. Roelens, *Angew. Chem., Int. Ed.*, 2006, **45**, 6693; (*y*) C. Li, G. T. Wang, H. P. Yi, X. K. Jiang, Z. T. Li and R. X. Wang, *Org. Lett.*, 2007, **9**, 1797; (*z*) C. Nativi, M. Cacciarini, O. Francesconi, A. Vacca, G. Moneti, A. Ienco and S. Roelens, *J. Am. Chem. Soc.*, 2007, **129**, 4377; M. Mazik and M. Kuschel, *Chem. Eur. J.*, 2008, **14**, 2405; (*aa*) H. Abe, H. Machiguchi, S. Matsumoto and M. Inouye, *J. Org. Chem.*, 2008, **73**, 4650; (*ab*) M. Mazik and A. Hartmann, *J. Org. Chem.*, 2008, **73**, 7444; (*ac*) P. B. Palde, P. C. Gareiss and B. L. Miller, *J. Am. Chem. Soc.*, 2008, **130**, 9566.
12 K. M. Bhattarai, A. P. Davis, J. J. Perry, C. J. Walter, S. Menzer and D. J. Williams, *J. Org. Chem.*, 1997, **62**, 8463.
13 A. P. Davis, *Chem. Soc. Rev.*, 1993, **22**, 243; A. P. Davis, R. P. Bonar-Law and J. K. M. Sanders, in *Comprehensive Supramolecular Chemistry*, ed. Y. Murakami, Pergamon, Oxford, 1996, Vol. 4 (Supramolecular Reactivity and Transport: Bioorganic Systems), p. 257.
14 E. A. Meyer, R. K. Castellano and F. Diederich, *Angew. Chem., Int. Ed.*, 2003, **42**, 1210.
15 N. K. Vyas, M. N. Vyas and F. A. Quiocho, *Science*, 1988, **242**, 1290.
16 A. P. Davis and R. S. Wareham, *Angew. Chem., Int. Ed.*, 1998, **37**, 2270.
17 Amides are stronger H-bond donors and acceptors than hydroxyl groups. See: M. H. Abraham and J. A. Platts, *J. Org. Chem.*, 2001, **66**, 3484.
18 A. P. Davis and R. S. Wareham, *Angew. Chem., Int. Ed.*, 1998, **37**, 2270.
19 T. J. Ryan, G. Lecollinet, T. Velasco and A. P. Davis, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 4863.
20 G. Lecollinet, A. P. Dominey, T. Velasco and A. P. Davis, *Angew. Chem., Int. Ed.*, 2002, **41**, 4093.
21 Selected examples: (*a*) K. Kobayashi, Y. Asakawa, Y. Kato and Y. Aoyama, *J. Am. Chem. Soc.*, 1992, **114**, 10307; (*b*) R. Yanagihara and Y. Aoyama, *Tetrahedron Lett.*, 1994, **35**, 9725; (*c*) B.-L. Poh and C. M. Tan, *Tetrahedron*, 1993, **49**, 9581; (*d*) H. Goto, Y. Furusho and E. Yashima, *J. Am. Chem. Soc.*, 2007, **129**, 9168; (*e*) Y. Aoyama, Y. Nagai, J. Otsuki, K. Kobayashi and H. Toi, *Angew. Chem., Int. Ed. Engl.*, 1992, **31**, 745; (*f*) A. V. Eliseev and H.-J. Schneider, *J. Am. Chem. Soc.*, 1994, **116**, 6081; (*g*) V. Král, O. Rusin and F. P. Schmidtchen, *Org. Lett.*, 2001,

3, 873; (*h*) O. Rusin, K. Lang and V. Král, *Chem. Eur. J.*, 2002, **8**, 655; (*i*) S. Striegler and M. Dittel, *J. Am. Chem. Soc.*, 2003, **125**, 11518; (*j*) M. Mazik and H. Cavga, *J. Org. Chem.*, 2006, **71**, 2957; (*k*) J. Billing, H. Grundberg and U. J. Nilsson, *Supramolecular Chemistry*, 2002, **14**, 367; (*l*) T. Reenberg, N. Nyberg, J. O. Duus, J. L. J. van Dongen and M. Meldal, *Eur. J. Org. Chem.*, 2007, 5003.
22 For more discussion, see ref. 8c.
23 T. Ryan, PhD thesis, University of Dublin, 2001.
24 E. Klein, M. P. Crump and A. P. Davis, *Angew. Chem., Int. Ed.*, 2005, **44**, 298.
25 Y. Ferrand, E. Klein, N. P. Barwell, M. P. Crump, J. Jiménez-Barbero, C. Vicent, G.-J. Boons, S. Ingale and A. P. Davis, *Angew. Chem., Int. Ed.*, 2009, **49**, 1775.
26 N. E. Zachara and G. W. Hart, *Chem. Rev.*, 2002, **102**, 431; G. W. Hart, M. P. Housley and C. Slawson, *Nature*, 2007, **446**, 1017.
27 S. Kubik, *Angew. Chem., Int. Ed.*, 2009, **48**, 1722.
28 Y. Ferrand, M. P. Crump and A. P. Davis, *Science*, 2007, **318**, 619.
29 E. J. Toone, *Curr. Opin. Struct. Biol.*, 1994, **4**, 719.
30 See, for example: E. García-Hernández and A. Hernández-Arana, *Protein Science*, 1999, **8**, 1075.
31 R. U. Lemieux, in *Carbohydrate Antigens*, ed. P. J. Garegg and A. A. Lindberg, American Chemical Society, 1993, p. 5; R. U. Lemieux, *Acc. Chem. Res.*, 1996, **29**, 373.
32 E. Klein, Y. Ferrand, N. P. Barwell and A. P. Davis, *Angew. Chem., Int. Ed.*, 2008, **48**, 2693.
33 M. Mazik, *ChemBioChem*, 2008, **9**, 1015.

