Carbohydrate Recognition

DOI: 10.1002/anie.200704733

Solvent Effects in Carbohydrate Binding by Synthetic Receptors: Implications for the Role of Water in Natural Carbohydrate Recognition**

*Emmanuel Klein, Yann Ferrand, Nicholas P. Barwell, and Anthony P. Davis**

Carbohydrate recognition is a key natural phenomenon[1] that mediates protein trafficking[1g] and function,[1h] cell–cell recognition and adhesion,[1i,j] and many aspects of the immune response. Despite its importance, it is not well understood in all respects. In particular, the driving force for saccharide binding by lectins and other carbohydrate-binding proteins is subject to debate.[2] The discussion centers on the part played by water. Crystal structures of protein–carbohydrate complexes reveal dense networks of intermolecular hydrogen bonds, but these bonds can only form after desolvation of the binding surfaces. Complex formation [Eq. (1)] involves the interchange of carbohydrate–OH groups with H_2O molecules, which is not an obviously favorable process.

$$\text{Protein}\cdot n\,H_2O + \text{Carbohydrate}\cdot m\,H_2O \rightleftharpoons \text{Complex}\cdot q\,H_2O + (n+m-q)\,H_2O \quad (1)$$

In view of Equation (1), it is not surprising that protein–carbohydrate interactions are relatively weak compared to other biomolecular associations.[2e,3] Also, it is unclear how natural saccharide binding is achieved. Broadly speaking, there are two hypotheses. The first suggests that the interaction is essentially polar with binding being driven by an especially favorable set of hydrogen bonds in the protein–carbohydrate complex.[4] The second proposes that solvent reorganization is a major player. Both carbohydrates and their complementary binding sites are amphiphilic with hydrophobic regions interspersed among the polar groups. It has been argued, especially by Lemieux,[2d,f] that their surfaces may be inefficiently hydrated so that the water on the left hand side of Equation (1) is high energy. Release of this water into bulk lowers the free energy and drives the formation of the complex.

Elegant methods developed for studying this problem involve thermochemistry,[5] calculational approaches,[2c] and mass spectrometry.[6] However, they focus mainly on enthalpies, as opposed to free energies, and they aim to determine the global influence of solvation (i.e. the overall difference between binding in the gas phase and solution). For a more complete picture one would like to examine the role of the solvent more closely, so that both negative (counterproductive) and positive contributions can be discerned.

A straightforward method for probing solvent effects in molecular recognition is to vary the solvent.[7] If the interactions between binding partners are exclusively polar, then the effect of polar solvent molecules will be uniformly negative. Increasing solvent polarity will lower the binding constants, and water, the most competitive medium for hydrogen-bonding interactions, will give the lowest affinities. If, however, there is a solvophobic contribution to the binding, the response may be different. Starting with nonpolar solvents, increases in solvent polarity will lower affinities as before, but when water is reached its exceptional cohesive properties may take effect. Affinities in pure water may therefore be higher than those in some less polar solvent systems.[8] Thus, the method can reveal a solvophobic component when polar interactions are also important.

Such studies may be problematic with natural receptors because, as proteins, they can denature (or at least change their conformations) when dissolved in nonaqueous media.[9] However, synthetic analogues[10] can be designed or manipulated such that they are compatible with a variety of solvent systems. We have developed a family of carbohydrate receptors exemplified by **1**[11] and **3**[12] (Figure 1). The macropolycyclic cores provide both polar (isophthalamide) and apolar (biphenyl or terphenyl) units for interacting with saccharides. The parallel disposition of the apolar units favors substrates with an all-equatorial arrangement of polar functional groups; for example, β-glucosyl **2** is suitable for **1** (see Figure 1a) and β-cellobioside **4** is compatible with the larger macrocycle (**3**). The architecture was inspired by the *E. coli* chemoreceptor protein in which the aromatic units are known to sandwich carbohydrate substrates (see Figure 1b).[10,13] Peripheral substituents are used to control solubilities so that both organic soluble (**1a** and **3a**) and water soluble (**1b** and **3b**) versions are available. The water soluble variants may be viewed as lectin models, which can bind to carbohydrates under natural conditions (i.e. in water) and show good to excellent selectivity for their target substrates. The binding of carbohydrates to **1b** was somewhat weak (K_a values up to $27\,M^{-1}$),[11] but carbohydrate binding to **3b** was more impressive (K_a values up to $900\,M^{-1}$, selectivities of approximately 50:1 for pairs of disaccharide substrates).[12]

[*] Dr. E. Klein, Dr. Y. Ferrand, N. P. Barwell, Prof. A. P. Davis
School of Chemistry
University of Bristol
Cantock's Close, Bristol, BS8 1TS (UK)
Fax: (+44) 117-929-8611
E-mail: anthony.davis@bris.ac.uk

[**] Support from the EU (RTN contract HPRN-CT-2002-00190) and EPSRC (grant EP/D060192/1) is gratefully acknowledged.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

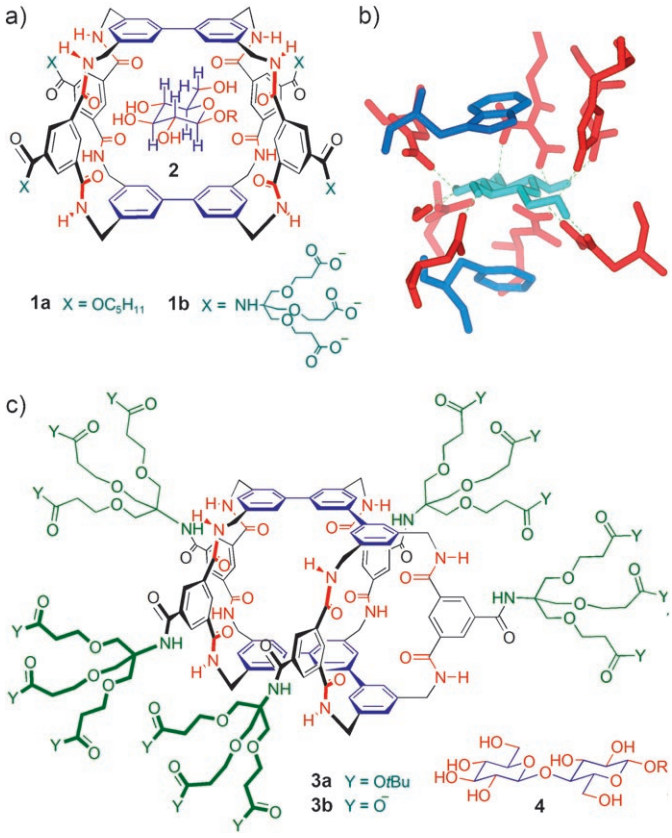

Figure 1. a) Monosaccharide receptors **1** enclosing β-glucosyl substrates **2**. Parallel biphenyl units make apolar contacts with axial CH groups (blue), whereas equatorial OR/OH groups form hydrogen bonds to the amide groups (red). Solubilizing groups are shown in green. b) Binding site of *E. coli* galactose chemoreceptor protein with a molecule of glucose substrate (cyan).[12] Aromatic tryptophan and phenylalanine residues are shown in blue and polar residues in red. c) Disaccharide receptors **3** and cellobiose substrate **4** with color coding as for a).

These biomimetic receptors presented an opportunity to study carbohydrate binding to reference systems in solvents with a wide range of polarities. The association constants of glucosyl substrates (**2**) to **1** had previously been measured in chloroform ($Z = 63.2$),[14] choloroform/methanol mixtures, and water ($Z = 94.6$) as illustrated in Table 1. The figures showed that, as expected for the recognition of polar molecules, affinities decrease as the solvent polarity increases. However, the data were too sparse to distinguish between exclusively polar interactions (monotonic decrease of the K_a value with increasing solvent polarity) and polar and solvophobic effects (the K_a value passes through a minimum). We therefore wanted to probe the solvent polarity space at a higher resolution, especially the critical region closer to pure water.

Table 1: The effect of widely differing media on the association constants (K_a) for the binding of glucosyl substrates **2** to receptors **1**.

Receptor	Substrate	Solvent	K_a [M^{-1}]
1a	**2** (R = C_8H_{17})	$CHCl_3$	300 000[a,b]
1a	**2** (R = C_8H_{17})	$CDCl_3/CD_3OH$ (92:8)	980[c,b]
1b	glucose[d]	D_2O	9.2[c,e]

[a] Determined by fluorescence titration. [b] See reference [11a]. [c] Determined by ^{1}H NMR titration. [d] α/β = 40:60. [e] See reference [11b].

The binding of **1b** to glucose and the binding of **3b** to cellobiose (**4**, R = H) were studied in a series of water/organic solvent mixtures in which the organic solvent was either MeOH ($Z = 83.6$), DMSO ($Z = 71.1$), or MeCN ($Z = 71.3$).[16] The proportion of the organic solvent was increased until the experiments became infeasible because of decreased solubility of the carbohydrate substrates. In the case of **1b**, the association constants were measured by ^{1}H NMR titrations[15] by analyzing movements of the signal from the inward-directed isophthaloyl CH group. For most of the experiments, analysis of two other aromatic CH signals was possible and gave similar values. Although the K_a values were low, the fits of the theoretical equations to the experimental data points were excellent.[16] In most cases[17] the receptor was greater than 65% saturated at the end of each titration. The results and the observed trends are therefore reliable. In the case of **3b**, fluorescence titrations were used; the increase in fluorescence output of the terphenyl unit was observed when cellobiose was added. In previous work this method gave results that were in agreement with those from NMR, induced CD, and calorimetric titrations.[12] Again the fits of theoretical equations to the experimental data were excellent, implying good accuracy.

Aggregation of the substrates was a potential complicating factor, especially in solvent mixtures with higher proportions of an organic solvent. However, the ^{1}H NMR spectra of both glucose and cellobiose were essentially concentration independent in the solvent systems used and implied that self association was minimal. The NMR spectra also showed that the anomeric ratios remained constant at α/β ≈ 40:60, throughout the experiments.

To provide a complementary set of results in organic media, the binding of **3a** to octyl β-D-cellobioside (**4**, R = C_8H_{17}) was investigated in a series of $CHCl_3$/MeOH solvent mixtures by fluorescence titrations.[16] β-D-Cellobioside **4** was freely soluble, with concentration independent ^{1}H NMR spectra, in solvents ranging from methanol to $CHCl_3$/MeOH (92:8) mixtures. It was less soluble in $CHCl_3$, but dissolved sufficiently in $CHCl_3$/MeOH (99.5:0.5) for the binding constant measurements.

The results of the binding experiments are summarized in Figures 2 and 3. Figure 2 includes the data for both the association of **3b**+cellobiose and **3a+4**, covering the full polarity range from water to chloroform mediated by methanol. Despite a gap in the sequence because of the poor solubility of cellobiose in methanol, it is clear that the interaction between the β-cellobiosyl unit and the core of receptors **3** does pass through a minimum. The K_a value for the binding of **3b** to cellobiose decays from 560 M^{-1} in water to 30 M^{-1} in H_2O/MeOH (75:25) (Figure 2a). For the association of **3a+4**, the affinities rise from 11 M^{-1} in MeOH to an apparent value of 3×10^6 M^{-1} in MeOH/$CHCl_3$ (0.5:99.5) (Figure 2b). This value may be distorted by substrate self association,[18] but if correct it places **3a** among the more powerful carbohydrate receptors studied in organic media.[19]

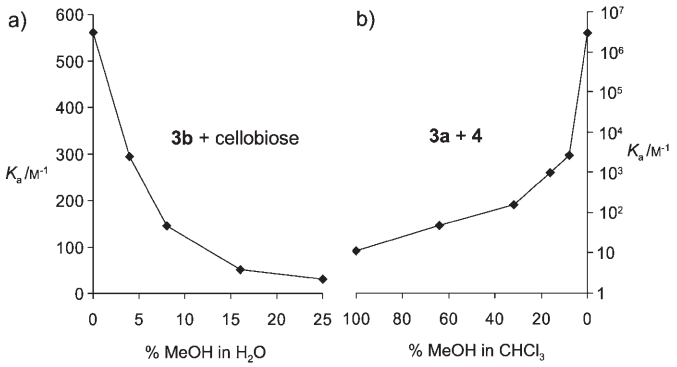

Figure 2. Binding constants (K_a) of disaccharide receptor **3** to cellobiosyl units in a series of water/methanol and methanol/chloroform solvent mixtures. a) Association constants for **3b**+D-cellobiose in H_2O/MeOH. b) Association constants for **3a**+octyl β-D-cellobioside (**4**, R = C_8H_{17}) in MeOH/$CHCl_3$. The K_a values in (b) are expressed on a logarithmic scale so that the full range of values can be represented.

The full set of results for aqueous/organic solvent mixtures shows that the effect of organic solvents on carbohydrate binding in water is a general phenomenon (Figure 3). It applies to both **1** and **3**, and to all three organic solvents tested. The degree of the effect depends on the organic solvent. MeCN is the most disruptive, next is DMSO, and then MeOH. The effects are moderate for the association of **1b**+glucose; 12% of added solvent lowered the K_a values by

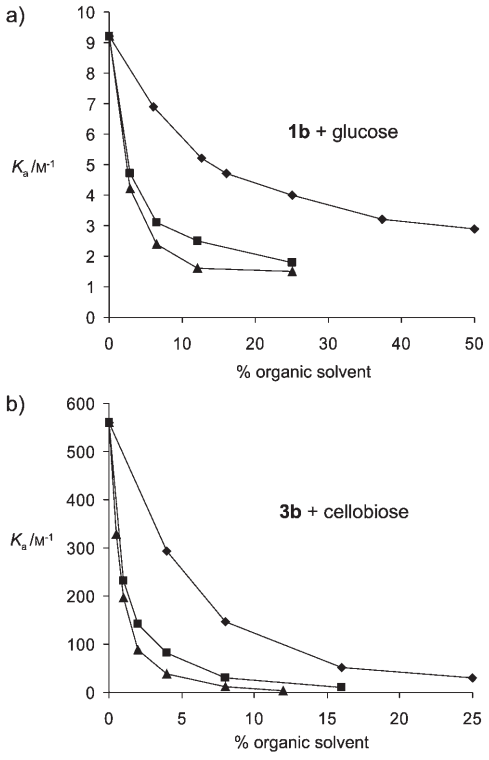

Figure 3. Binding constants (K_a) of synthetic receptors to carbohydrate substrates in water/organic solvent mixtures. Organic solvents: ◆ = MeOH, ■ = DMSO, ▲ = MeCN. a) Association constants for **1b**+glucose measured by ^{1}H NMR spectroscopy (deuterated solvents). b) Association constants for **3b**+cellobiose measured by fluorescence spectroscopy (data for MeOH is repeated from Figure 2a).

factors of 1.8, 3.7, and 5.8 for MeOH, DMSO, and MeCN, respectively. For the binding constants for **3b**+cellobiose the reductions are very substantial; for just 8% of added MeOH, DMSO, or MeCN, the factors were 3.8, 28, and 47, respectively. The difference between the two receptors presumably relates to their sizes. The extended apolar surfaces in **3b** and its cellobiose substrate result in larger hydrophobic effects, which are more sensitive to the disruptive effects of the organic solvent. The results may be compared with those of Diederich and co-workers on the hydrophobic binding of substituted benzene rings by a water soluble cyclophane.[20] In that case approximately 10% of either MeOH or DMSO caused a roughly twofold reduction in affinities. It is interesting that the same two solvents have different effects in the present case, and the systems are much more sensitive to DMSO. This sensitivity may relate to the role of both the polar and hydrophobic interactions in carbohydrate recognition.

As an additional test of the role of water, we investigated the effect KSCN. Such chaotropic salts are generally understood to weaken hydrophobic interactions,[21] and indeed, we observed lower affinities; for [KSCN] = 0.3 M, the K_a values for the associations of **1b**+glucose and **3b**+cellobiose fell by factors of 3 and 6.4, respectively.

Taken together, the results imply a major role for hydrophobic forces in carbohydrate recognition by this family of synthetic receptors. Although the classical hydrophobic effect is entropy-driven, it is notable that the binding of cellobiose by **3b** in water is largely enthalpy-driven ($\Delta H = -3.22$ kcal mol^{-1}; $T\Delta S = 0.62$ kcal mol^{-1}).[12] This may therefore be seen as another example of the hydraphobic[2d] or nonclassical hydrophobic[7a] effect.

The implications of our results for biological carbohydrate recognition depend on the degree to which **1b** and **3b** are truly biomimetic. The binding sites bear good general resemblances to those of carbohydrate-binding proteins, being amphiphilic in character with preorganized polar and apolar moieties. The designs are specifically related to carbohydrate-binding proteins, and NMR structural studies suggest that they work as predicted.[11b,12] Although the binding constants are quite low (especially for **1b**), recall that carbohydrate–protein interactions are also weak; lectin–monosaccharide affinities are often less than 10^3 M^{-1}.[2e] Importantly, the thermodynamic parameters for association of **3b**+cellobiose in water (see above) lie within the range observed for lectins.[2e] These considerations suggest that **1b** and **3b** are indeed good models for natural carbohydrate receptors and can be used to shed light onto their biological counterparts. Whereas saccharide binding in nature is strongly influenced by hydrogen bonding, our data provide clear support for a hydrophobic component.

Received: October 12, 2007
Revised: January 3, 2008
Published online: February 22, 2008

Keywords: carbohydrates · hydrogen bonds · hydrophobic effect · molecular recognition · solvent effects

[1] Leading references: a) H. J. Gabius, H. C. Siebert, S. Andre, J. Jiménez-Barbero, H. Rudiger, *ChemBioChem* **2004**, *5*, 740; b) R. A. Dwek, T. D. Butters, *Chem. Rev.* **2002**, *102*, 283; c) C. R. Bertozzi, L. L. Kiessling, *Science* **2001**, *291*, 2357; d) S. J. Williams, G. J. Davies, *Trends Biotechnol.* **2001**, *19*, 356; e) T. Feizi, B. Mulloy, *Curr. Opin. Struct. Biol.* **2001**, *11*, 585; f) *Carbohydrates in Chemistry and Biology* (Eds.: B. Ernst, G. W. Hart, P. Sinay), Wiley-VCH, Weinheim, **2000**; g) J. Roth, *Chem. Rev.* **2002**, *102*, 285; h) N. E. Zachara, G. W. Hart, *Chem. Rev.* **2002**, *102*, 431; i) S.-i. Hakomori, K. Handa in *Carbohydrates in Chemistry and Biology* (Eds.: B. Ernst, G. W. Hart, P. Sinay), Wiley-VCH, Weinheim, **2000**, p. II 771; j) M. Aubery, C. Derappe in *Carbohydrates in Chemistry and Biology* (Eds.: B. Ernst, G. W. Hart, P. Sinay), Wiley-VCH, Weinheim, **2000**, p. II 909.

[2] a) M. Ambrosi, N. R. Cameron, B. G. Davis, *Org. Biomol. Chem.* **2005**, *3*, 1593; b) H. J. Gabius, *Naturwissenschaften* **2000**, *87*, 108; c) E. García-Hernández, A. Hernández-Arana, *Protein Sci.* **1999**, *8*, 1075; d) R. U. Lemieux, *Acc. Chem. Res.* **1996**, *29*, 373; e) E. J. Toone, *Curr. Opin. Struct. Biol.* **1994**, *4*, 719; f) R. U. Lemieux in *Carbohydrate Antigens* (Eds.: P. J. Garegg, A. A. Lindberg), American Chemical Society, New York, **1993**, p. 5; g) F. A. Quiocho, *Pure Appl. Chem.* **1989**, *61*, 1293; h) M. C. Fernandez-Alonso, F. J. Canada, J. Jiménez-Barbero, G. Cuevas, *J. Am. Chem. Soc.* **2005**, *127*, 7379.

[3] J. J. Lundquist, E. J. Toone, *Chem. Rev.* **2002**, *102*, 555.

[4] For a clear statement of this position, see reference [2c].

[5] M. C. Chervenak, E. J. Toone, *J. Am. Chem. Soc.* **1994**, *116*, 10533.

[6] E. N. Kitova, D. R. Bundle, J. S. Klassen, *Angew. Chem.* **2004**, *116*, 4279; *Angew. Chem. Int. Ed.* **2004**, *43*, 4183.

[7] Leading references: a) E. A. Meyer, R. K. Castellano, F. Diederich, *Angew. Chem.* **2003**, *115*, 1244; *Angew. Chem. Int. Ed.* **2003**, *42*, 1210; b) G. A. Breault, C. A. Hunter, P. C. Mayers, *J. Am. Chem. Soc.* **1998**, *120*, 3402.

[8] For a relevant example involving aromatic interactions, see reference [7b].

[9] The lectin hevein has been studied in [D_6]DMSO/water mixtures; see: H. C. Siebert, S. Andre, J. L. Asensio, F. J. Canada, X. Dong, J. F. Espinosa, M. Frank, M. Gilleron, H. Kaltner, T. Kozar, N. V. Bovin, C. W. von der Lieth, J. F. G. Vliegenthart, J. Jiménez-Barbero, H. J. Gabius, *ChemBioChem* **2000**, *1*, 181. The results were consistent with, and complementary to, the data reported in this paper.

[10] a) A. P. Davis, R. S. Wareham, *Angew. Chem.* **1999**, *111*, 3160; *Angew. Chem. Int. Ed.* **1999**, *38*, 2978; b) A. P. Davis, T. D. James in *Functional Synthetic Receptors* (Eds.: T. Schrader, A. D. Hamilton), Wiley-VCH, Weinheim, **2005**, p. 45.

[11] a) A. P. Davis, R. S. Wareham, *Angew. Chem.* **1998**, *110*, 2397; *Angew. Chem. Int. Ed.* **1998**, *37*, 2270; b) E. Klein, M. P. Crump, A. P. Davis, *Angew. Chem.* **2005**, *117*, 302; *Angew. Chem. Int. Ed.* **2005**, *44*, 298.

[12] Y. Ferrand, M. P. Crump, A. P. Davis, *Science* **2007**, *318*, 619.

[13] N. K. Vyas, M. N. Vyas, F. A. Quiocho, *Science* **1988**, *242*, 1290.

[14] Kosower's *Z* parameter is used to quantify solvent polarity. See: E. M. Kosower, *J. Am. Chem. Soc.* **1958**, *80*, 3253. Units are kcal mol^{-1} and the values are taken from reference [7b].

[15] Solvents were deuterated for these experiments.

[16] For further details, see the Supporting Information.

[17] Exceptions are the three examples where $K_a < 2\,\text{M}^{-1}$. These results (only) are therefore tentative.

[18] Aggregation of the cellobioside cannot be discounted in this nonpolar solvent system. However, the risk is minimized by the low substrate concentrations required for the binding experiment (less than 0.7 μM). If substrate self association did occur it would decrease the amount of free glycoside available for binding, tending to cause an underestimate of the binding constant.

[19] For binding in chloroform, values up to $2\times10^7\,\text{M}^{-1}$ have been reported. See reference [10b] and Y. H. Kim, J. I. Hong, *Angew. Chem.* **2002**, *114*, 3071; *Angew. Chem. Int. Ed.* **2002**, *41*, 2947.

[20] S. B. Ferguson, E. M. Sanford, E. M. Seward, F. Diederich, *J. Am. Chem. Soc.* **1991**, *113*, 5410.

[21] K. A. Dill, T. M. Truskett, V. Vlachy, B. Hribar-Lee, *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 173.

