Carbohydrate Recognition

DOI: 10.1002/ange.201200447

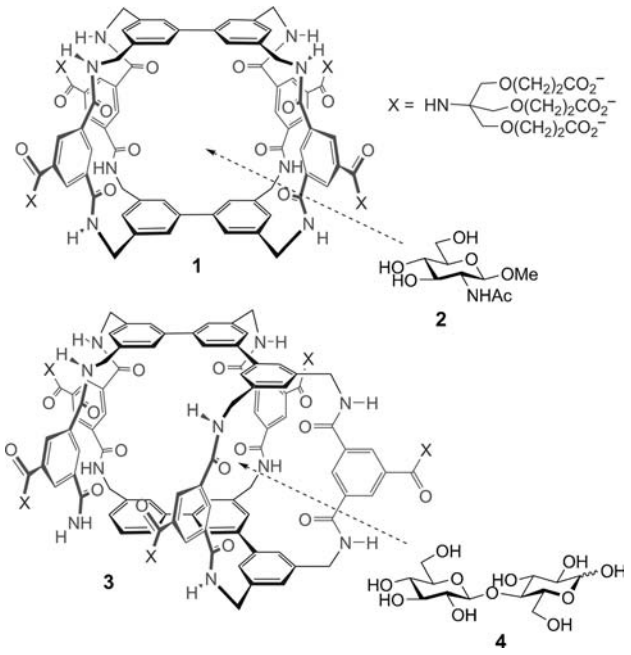
High-Affinity Disaccharide Binding by Tricyclic Synthetic Lectins**

*Bunyarithi Sookcharoenpinyo, Emmanuel Klein, Yann Ferrand, D. Barney Walker, Peter R. Brotherhood, Chenfeng Ke, Matthew P. Crump, and Anthony P. Davis**

Carbohydrate recognition mediates a wide range of biological processes,[1] including protein folding and trafficking,[2] cell–cell recognition,[3] infection by pathogens,[4] and many aspects of the immune response.[5] Molecules capable of selective carbohydrate binding are, therefore, valuable as tools for biological research, and potentially as medicinal agents.[6,7] There are many saccharide-binding proteins, notably the group known as lectins,[8] but they often show low affinities and (from a researcher's viewpoint) non-ideal selectivities.[8,9] Moreover, they are generally too unstable and toxic for use in medicine. Opportunities thus exist for synthetic systems,[10] provided they can compete with lectins in terms of binding strength and selectivity.

Over the past few years we have shown that macropolycyclic lactams such as **1** and **3** (Scheme 1) can bind carbohydrates with all-equatorial substitution patterns in water, by employing a combination of hydrogen bonding and hydrophobic/CH-π interactions.[11] Selectivities compare well with those of lectins; for example, both **1** and **3** bind their respective targets **2** and **4** with preferences of $>20{:}1$ versus closely related substrates.[11b,c] However, affinities have been less competitive. While lectins typically bind monosaccharides and disaccharides with $K_a = 10^3$–$10^5\,\text{M}^{-1}$,[8,9] the values for **1** and **3**, at approximately $600\,\text{M}^{-1}$, lie short of this range. We now report two new synthetic lectins **5** and **6**, which are related to tetracycle **3** but with a less preorganized and more accessible architecture. Remarkably, these systems yield increased performance in key respects, with K_a values up to $4500\,\text{M}^{-1}$ and extreme selectivity for di- versus monosaccharides. This study provides the best evidence yet that true lectin mimicry may be achieved with simple, practically viable receptor structures.

The decision to undertake this study was made by an indirect route, which nicely illustrates the limitations of current molecular design. When initially we planned to target all-equatorial disaccharides, tricyclic system **5** seemed an obvious solution.[12] However, modeling studies[13] showed that **5** can undergo a twisting motion which brings together the central aromatic rings in each terphenyl unit (Figure 1). In water, this collapsed conformation was predicted to be $\geq 25\,\text{kJ mol}^{-1}$ more stable than open structures, presumably compromising disaccharide binding. We therefore turned to the more highly connected **3** which, as previously described,[11b] proved very effective.



Scheme 1. Previously reported macropolylactam synthetic lectins **1** and **3**, with favored substrates D-GlcNAc-β-OMe (**2**) and D-cellobiose (**4**). Binding is driven by hydrogen bonding to equatorial polar groups in the carbohydrates, and hydrophobic/CH-π interactions to an axial CH group of the substrate.

Despite the success of **3**, we remained curious whether the rigid tetracyclic architecture was necessary or, alternatively, the tricyclic (and synthetically more accessible) **5** might show at least moderate binding properties. Having recently found that 4,4′-alkoxybiphenyl substituents could improve affinities in monosaccharide receptors,[11d] we were also interested to test the dimethoxy analogue **6**. We therefore synthesized **5** and **6** from haloarenes **7** by sequential Suzuki–Miyaura cross-coupling reactions, deprotections, and macrolactamizations (Scheme 2).[14] The method was largely based on earlier work, but a new development was the use of diazido aryl halides **8** as coupling partners for boronates **9**. While care was required to prevent azide degradation during coupling, the Staudinger reduction of tetraazido macrocycles **11** proceeded especially smoothly. This protection strategy proved superior to alternatives such as benzyloxycarbonyl (Cbz), for which removal could be problematic.

[*] B. Sookcharoenpinyo, Dr. E. Klein, Dr. Y. Ferrand, Dr. D. B. Walker, Dr. P. R. Brotherhood, Dr. C. Ke, Dr. M. P. Crump, Prof. A. P. Davis
School of Chemistry, University of Bristol
Cantock's Close, Bristol, BS8 1TS (UK)
E-mail: anthony.davis@bristol.ac.uk

[**] This work was supported by the Royal Thai Government (Fellowship to B.S.), the Engineering and Physical Sciences Research Council (grant number EP/D060192/1), and the Royal Society (Newton International Fellowship to C.K.)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201200447.

© 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

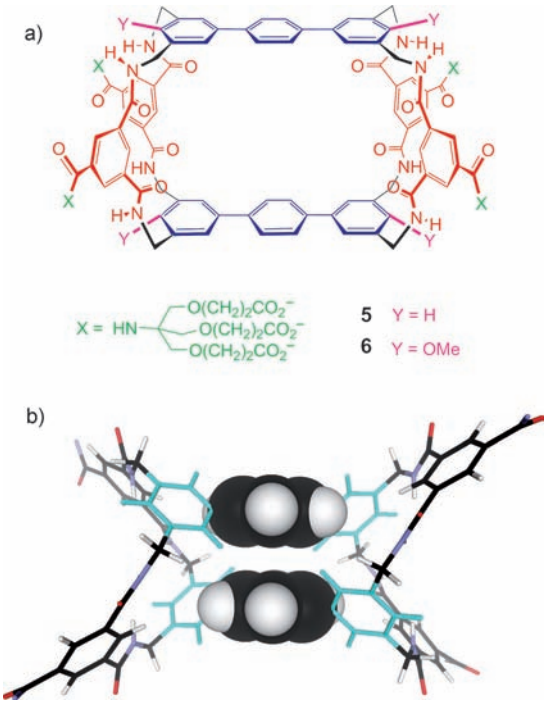
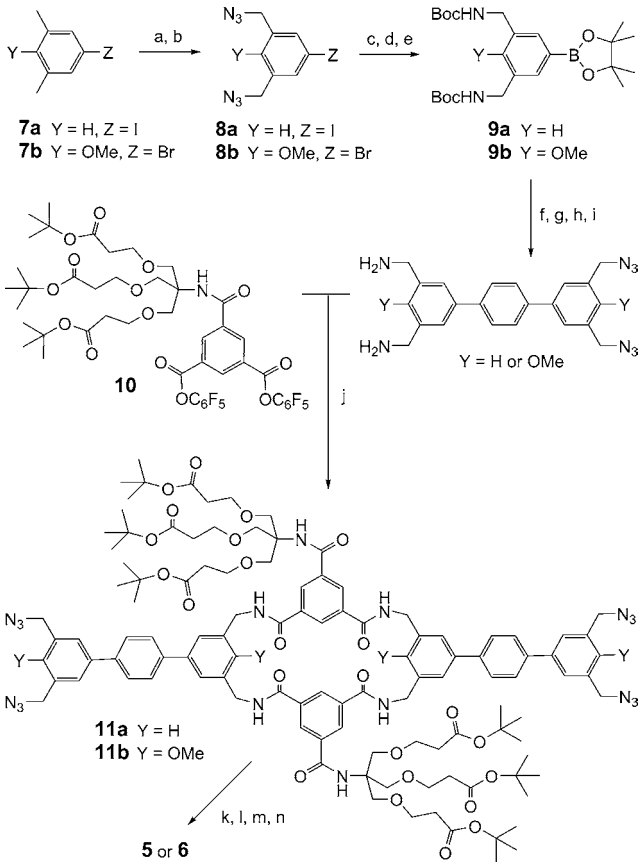


Figure 1. a) Formulas for receptors **5** and **6**. b) Ground-state conformation of **5** as predicted by Monte Carlo Molecular Mechanics (MCMM).[13,14] The two ends of the molecule have twisted relative to each other, bringing the central terphenyl aromatic rings into proximity. The central rings are shown in space-filling mode, and the terminal terphenyl aromatic rings are highlighted in cyan. Water-solubilizing groups are omitted for clarity.

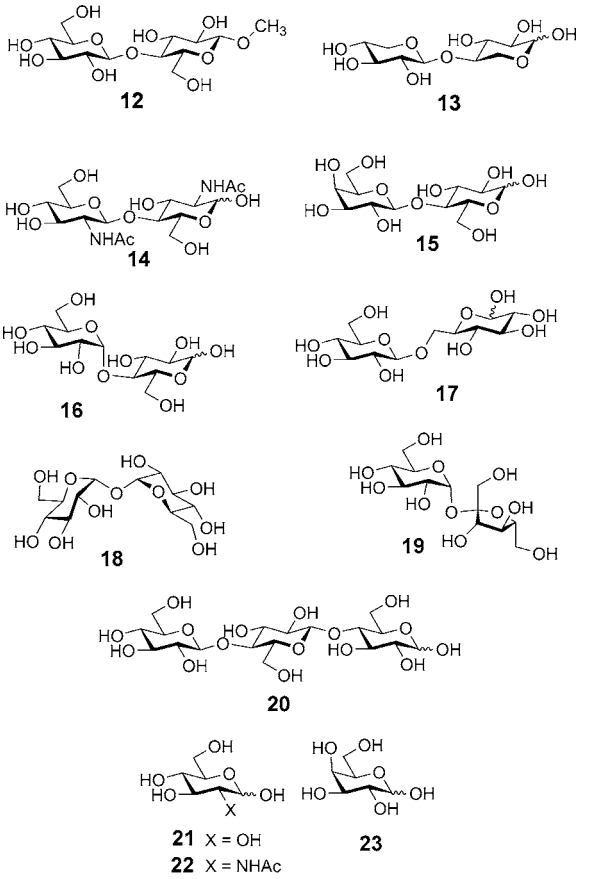


Scheme 2. Synthesis of receptors **5** and **6**. Yields refer to Y = H/Y = OMe, respectively: a) *N*-bromosuccinimide, benzoyl peroxide (cat.), MeOAc, 85 °C, 48 h, 40/99%; b) NaN_3, DMF, 60 °C, 24 h, ca. 99%; c) PPh_3, THF/H_2O, 60 °C, 24 h, 95/99%; d) Boc_2O, DIPEA, THF, RT, 24 h, 75/83%; e) pin_2B_2, $[PdCl_2(dppf)]$ (3 mol%), KOAc, DMSO, 80 °C, 18 h, 80/83%; f) 4-bromo-iodobenzene, $[PdCl_2(dppf)]$ (3 mol%), Na_2CO_3, DMSO, 50 °C, 24 h, 85%; g) pin_2B_2, $[PdCl_2(dppf)]$ (3 mol%), KOAc, DMSO, 80 °C, 2 days, 85/94%; h) bisazide **8a** or **8b**, $[PdCl_2(dppf)]$ (3 mol%), Na_2CO_3, DMF, 80 °C, 4 h, 72/74%; i) TFA, CH_2Cl_2, 0 °C, 5 h, 78/89%; j) DIPEA, THF, 30 h injection, RT, 24 h, 20/52%; k) PPh_3, THF/H_2O, 60 °C, 24 h, 78/91%; l) **10**, THF, 30 h injection, RT, 24 h, 29/58%; m) TFA, CH_2Cl_2, 0 °C, 5 h, 98/99%; n) aq NaOH, 95/98%. Boc = *tert*-butoxycarbonyl, DIPEA = *N*,*N*-diisopropylethylamine, pin_2B_2 = bispinacolato diboron, $[PdCl_2(dppf)]$ = dichlorodiphenylphosphinoferrocene palladium(II) complex, TFA = trifluoroacetic acid.

Macrocycles **5** and **6** were examined first by ^{1}H NMR spectroscopy in D_2O. The parent system **5** gave broadened signals which could not be used for binding studies, but fortunately the tetramethoxy analogue **6** gave well-resolved spectra. Dilution of an NMR sample from 0.5 mM to 0.15 mM caused negligible changes to the spectra, thus implying that **6** does not aggregate in this concentration range. Carbohydrate recognition by **6** was then studied by ^{1}H NMR titrations, using cellobiose (**4**) and saccharides **12**–**23** (Scheme 3) as substrates. The addition of **4** as well as several other substrates caused movements of signals in the aromatic region of the spectrum, which implies binding with fast-medium exchange on the ^{1}H NMR timescale (Figure 2). Broadening was observed for some signals in some cases, but at least one proton could always be followed easily throughout a titration. The movements gave excellent fits to a 1:1 binding model, and were analyzed to give the binding constants K_a listed in Table 1. Interestingly, the signals for aromatic protons *A* and *C* were split in some cases, while those for *B* and *D* remained singlets. This is consistent with the D_{2h} symmetry and conformational properties of the host. A carbohydrate entering the host renders all the protons different but, when fast exchange occurs, movement between four substrate orientations results in equivalence for groups of four protons. Thus the signal for *B* (4 protons) remains a singlet, while those of *A* (8 protons) and *C* (8 protons) are split into two groups of 4. The 8 protons of *D* appear as a singlet, because the two environments generated by the carbohydrate are exchanged by rotation of the C_6H_4 unit.

To support the NMR studies on **6**, we also employed fluorescence titrations and isothermal titration calorimetry (ITC). The addition of carbohydrates to a solution of **6** in water caused increases in the fluorescence output which, though modest, fit well to a 1:1 binding model (for example, see Figure 3). The K_a values obtained by this method are compared to the NMR-derived figures in Table 1. The ITC traces (e.g. Figure 3) were also consistent with 1:1 complexation,[14] and gave the thermodynamic parameters shown in Table 2. The fluorescence and ITC techniques were also applicable to parent tricycle **5**. Receptor **5** behaved similarly to **6** in these experiments, and gave the binding constants

 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim



Scheme 3. Carbohydrates used as substrates for binding studies.

Table 1: Association constants (K_a) for 1:1 complexes of receptors **5** and **6** with carbohydrate substrates in water, as determined by ^{1}H NMR and fluorescence titrations. Data for **1** and **3** are given for comparison.[a]

Carbohydrate	K_a [M^{-1}]				
	1	**3**	**5**[b]	**6**[c]	**6**[b]
methyl β-D-cellobioside (**12**)		850		4500	4470
D-cellobiose (**4**)	17	580	3140	3340	3330
D-xylobiose (**13**)		270		210	
N,*N*′-diacetyl-D-chitobiose (**14**)	ca. 0	120		910	
D-lactose (**15**)	ca. 0	14	230	270	240
D-maltose (**16**)	ca. 0	11	67	72	80
D-gentiobiose (**17**)		5		35	36
D-trehalose (**18**)		ca. 0		ca. 0	ca. 0
D-sucrose (**19**)		ca. 0		ca. 0	ca. 0
D-cellotriose (**20**)				87	
D-glucose (**21**)	9	11[d]		2	3
N-acetyl-D-glucosamine (**22**)	56	24		9	9
D-galactose (**23**)	2			ca. 0	

[a] $T = 298$ K unless otherwise stated. Values for **1** and **3** from Refs. [11a, b], respectively. Values denoted ca. 0 were too small for analysis. Values for reducing sugars are weighted averages of those for the two anomers, as discussed in Ref. [11d]. [b] Fluorescence titration in H_2O. Errors estimated at $\leq 4\%$. [c] ^{1}H NMR titration in D_2O. Errors estimated at $\leq 5\%$. [d] $T = 278$ K.

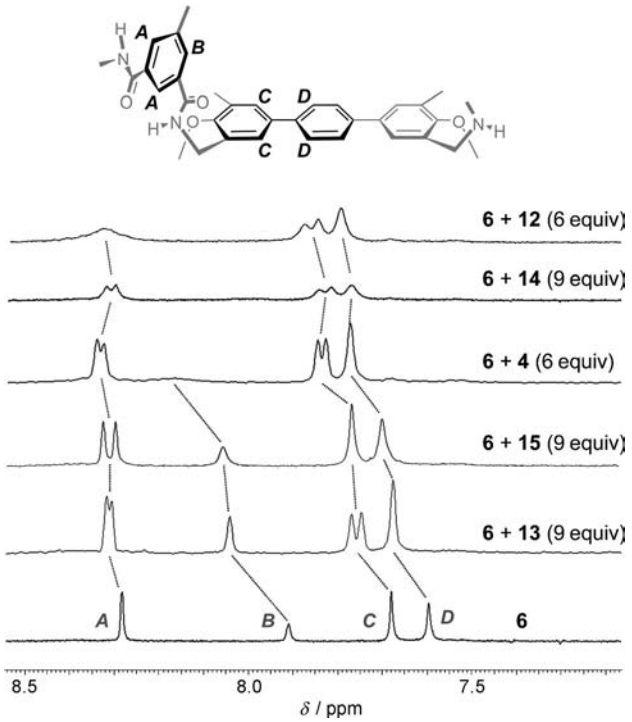


Figure 2. Assigned partial ^{1}H NMR spectra (500 MHz, D_2O, 298 K) of **6** alone (0.5 mM) and with added methyl β-D-cellobioside (**12**), *N*,*N*′-diacetyl-D-chitobiose (**14**), D-cellobiose (**4**), D-lactose (**15**), and D-xylobiose (**13**). Correlations between signals are indicated by dotted lines.

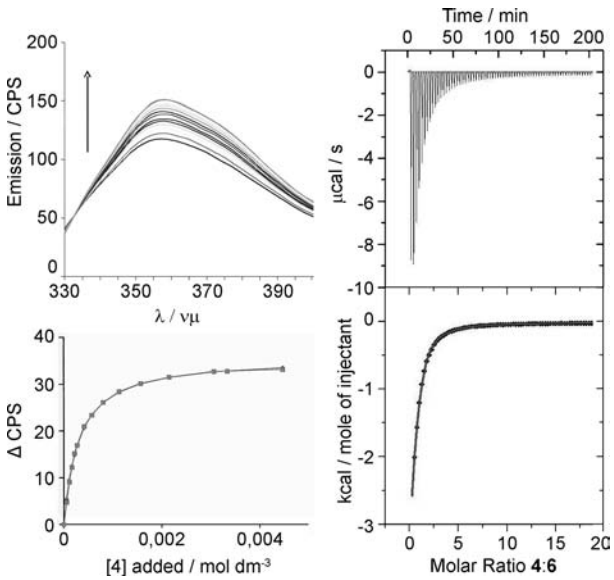


Figure 3. Data and analysis curves for (left) fluorescence and (right) ITC binding studies on receptor **6** + cellobiose **4** in H_2O at 298 K.[14] Both analyses employ a simple 1:1 binding model. $K_a = 3330$ and 3300 M^{-1}, respectively.

listed in Tables 1 and 2, respectively. For both **5** and **6**, the agreement between different methods was excellent, especially for stronger-binding substrates. For example, in the case of **6** + **4**, the NMR spectroscopic, fluorescence, and ITC measurements gave $K_a = 3340$, 3330, and 3300 M^{-1}, respectively.

Two features stand out when the binding constants in Tables 1 and 2 are considered. Firstly, against our expect-

© 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 2: Association constants (K_a) and thermodynamic quantities for 1:1 host–guest complexes between receptors **3**, **5** and **6** and carbohydrate substrates in water at 298 K, as measured by ITC.

Complex	K_a [M^{-1}]	ΔH [kJ mol^{-1}]	$T\Delta S$ [kJ mol^{-1}]	ΔG [kJ mol^{-1}]
3 + cellobiose (**4**)[a]	650	−13.46	2.59	−16.05
5 + cellobiose (**4**)	3110	−18.37	1.52	−19.89
5 + lactose (**15**)	220	−8.41	4.90	−13.30
5 + maltose (**14**)	61	−3.60	6.82	−10.42
6 + cellobiose (**4**)	3300	−18.86	1.23	−20.08
6 + lactose (**15**)	250	−9.82	3.82	−13.64
6 + maltose (**14**)	89	−4.45	6.39	−10.83

[a] From Ref. [11b].

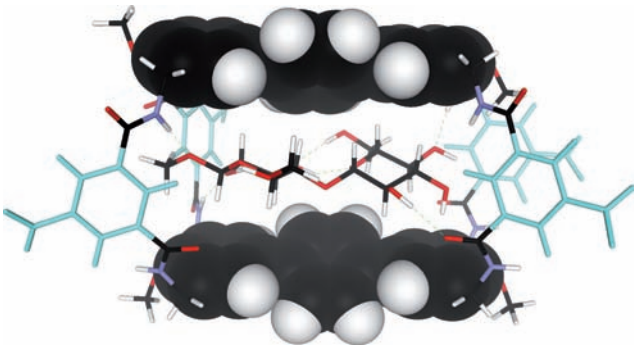


Figure 4. Possible structure for complex **6**·**12** derived from molecular modeling studies. Shown is the ground-state geometry from an MCMM search in which both the receptor and substrate were allowed conformational freedom.[14] Terphenyl aromatic groups are shown in space-filling mode, isophthalamide aromatic rings are colored cyan. Water-solubilizing groups are omitted for clarity.

ations, the affinities of both **5** and **6** for their target substrates **4** and **12** are considerably higher than achieved by the more rigid system **3**. Removing the fifth isophthalamide bridge, straightening the terphenyl units, and allowing the framework to flex has strengthened binding by factors of 5–6. Indeed, the affinity of **6** for methyl cellobioside (**12**), at 4500 M^{-1}, is the highest yet observed for a synthetic receptor binding an uncharged carbohydrate substrate in water. Perhaps surprisingly, given our previous experience with monosaccharide receptors,[11d] the methoxy substituents made little difference; the binding constants to **5** and **6** were almost identical. Secondly, the selectivity between disaccharides has been reduced (e.g. cellobiose/lactose = 40:1 for **3**, 13:1 for **6**), but the preference for disaccharides versus monosaccharides has been increased. Thus the cellobiose/glucose selectivity has risen from 50:1 for **3** to about 1300:1 for **6**. Selectivity between di- and monosaccharides may not be especially difficult to achieve, but again this system sets new records.

An advantage of the generally high affinities is that accurate ITC measurements are feasible for **5** and **6** with both target and nontarget substrates. As shown in Table 2, binding is driven by both enthalpy and entropy changes, but with enthalpy–entropy compensation such that enthalpy dominates for strongly bound complexes. Comparison with literature data[9] shows that these numbers are within the bounds observed for lectins, although with entropy contributions which are above-average for the natural systems. An enthalpy–entropy plot based on Table 2 is almost linear (see Figure S69 in the Supporting Information), as observed for other closely related host–guest pairings.[15]

Structural aspects of binding were investigated by NMR spectroscopy and molecular modeling studies. We focused especially on **6** + **12**, which form a strongly bound complex and avoid the complications caused by substrate anomers. Intermolecular cross-peaks in the NOESY spectrum provided clear evidence that, as expected, **12** enters the cavity of **6** to make hydrophobic/CH-π contacts with the terphenyl units. Thus, strong interactions were observed between axial substrate CH and receptor protons *C* and *D* (see Figure 2 for labeling), while cross-peaks from *C*/*D* to the C(6) protons were weaker. In particular, a correlation between the cellobioside MeOC*H* proton and one receptor proton *C* places the methyl glycosidic unit unambiguously in one corner of the cavity. MCMM calculations generated a number of conformations consistent with these data (Figure 4). Typically, these feature 4–6 intermolecular hydrogen bonds and approximately 10 CH-π interactions. Molecular dynamics simulations predicted that the complex should stay intact for at least 10 ns at 300 K.[14]

It is interesting to compare the structure in Figure 4 with the NOESY-based model previously obtained for **3**·**4**. Unsurprisingly, given the additional spacer unit in **3**, the latter complex features more intermolecular hydrogen bonds (ca. 10). On the other hand, the linear *p*-terphenyl unit in **6** may be slightly more compatible with the cellobiose CH group than the bent, *m*-substituted system in **3** (see Figures S74/75 in the Supporting Information). Thus, the strength of binding to **5** and **6** tends to reinforce our view that apolar hydrophobic/CH-π interactions provide the major driving force for carbohydrate recognition in water.[16] Whatever the reason, it is remarkable that **5** and **6** can outperform **3** despite the (at most) transient nature of their cavities (cf. Figure 1).

In conclusion, we have shown that tricyclic synthetic lectins **5** and **6** are even more effective than tetracyclic **3** at binding all-equatorial disaccharides under biomimetic conditions. The success of these less-connected structures suggests that an "induced fit" or "conformational selection"[17] approach can be superior to rigid preorganization in carbohydrate recognition, and may point the way to new, even simpler systems with potential for applications. In particular, the disaccharide substrates are representative of major biopolymers (cellulose, xylan, chitin), and receptors which bind the polymers themselves could have biological activity or serve as aids to processing (e.g. by promoting solubility). Future studies will focus on these possibilities.

Received: January 17, 2012
Published online: March 29, 2012

Keywords: biomimetic hosts · carbohydrates · molecular recognition · receptors · supramolecular chemistry

 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

[1] *The Sugar Code—Fundamentals of Glycoscience* (Ed.: H.-J. Gabius), Wiley-Blackwell, Weinheim, **2009**; H. J. Gabius, H. C. Siebert, S. Andre, J. Jiménez-Barbero, H. Rudiger, *ChemBioChem* **2004**, *5*, 740; R. A. Dwek, T. D. Butters, *Chem. Rev.* **2002**, *102*, 283; C. R. Bertozzi, L. L. Kiessling, *Science* **2001**, *291*, 2357.
[2] J. Roth, *Chem. Rev.* **2002**, *102*, 285.
[3] 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.
[4] T. Angata, A. Varki, *Chem. Rev.* **2002**, *102*, 439.
[5] R. Kannagi in *Carbohydrate-Based Drug Discovery* (Ed.: C.-H. Wong), Wiley-VCH, Weinheim, **2003**, p. 803.
[6] J. Balzarini, *Nat. Rev. Microbiol.* **2007**, *5*, 583.
[7] S. Jin, Y. F. Cheng, S. Reid, M. Y. Li, B. H. Wang, *Med. Res. Rev.* **2010**, *30*, 171.
[8] M. Ambrosi, N. R. Cameron, B. G. Davis, *Org. Biomol. Chem.* **2005**, *3*, 1593; H. Lis, N. Sharon, *Chem. Rev.* **1998**, *98*, 637.
[9] E. J. Toone, *Curr. Opin. Struct. Biol.* **1994**, *4*, 719.
[10] For reviews, see Ref. [7] and: D. B. Walker, G. Joshi, A. P. Davis, *Cell. Mol. Life Sci.* **2009**, *66*, 3177; M. Mazik, *Chem. Soc. Rev.* **2009**, *38*, 935; T. D. James, M. D. Phillips, S. Shinkai, *Boronic Acids in Saccharide Recognition*, RSC, Cambridge, **2006**; A. P. Davis, R. S. Wareham, *Angew. Chem.* **1999**, *111*, 3160; *Angew. Chem. Int. Ed.* **1999**, *38*, 2978; For recent examples see: G. Fukuhara, Y. Inoue, *J. Am. Chem. Soc.* **2011**, *133*, 768; A. Pal, M. Berube, D. G. Hall, *Angew. Chem.* **2010**, *122*, 1534; *Angew. Chem. Int. Ed.* **2010**, *49*, 1492; M. Rauschenberg, S. Bomke, U. Karst, B. J. Ravoo, *Angew. Chem.* **2010**, *122*, 7498; *Angew. Chem. Int. Ed.* **2010**, *49*, 7340; S. Striegler, M. G. Gichinga, *Chem. Commun.* **2008**, 5930; T. Reenberg, N. Nyberg, J. O. Duus, J. L. J. van Dongen, M. Meldal, *Eur. J. Org. Chem.* **2007**, 5003; H. Goto, Y. Furusho, E. Yashima, *J. Am. Chem. Soc.* **2007**, *129*, 9168; M. Mazik, H. Cavga, *J. Org. Chem.* **2006**, *71*, 2957.
[11] a) E. Klein, M. P. Crump, A. P. Davis, *Angew. Chem.* **2005**, *117*, 302; *Angew. Chem. Int. Ed.* **2005**, *44*, 298; b) Y. Ferrand, M. P. Crump, A. P. Davis, *Science* **2007**, *318*, 619; c) Y. Ferrand, E. Klein, N. P. Barwell, M. P. Crump, J. Jiménez-Barbéro, C. Vicent, G. J. Boons, S. Ingale, A. P. Davis, *Angew. Chem.* **2009**, *121*, 1807; *Angew. Chem. Int. Ed.* **2009**, *48*, 1775; d) N. P. Barwell, M. P. Crump, A. P. Davis, *Angew. Chem.* **2009**, *121*, 7809; *Angew. Chem. Int. Ed.* **2009**, *48*, 7673.
[12] We had previously found that a lipophilic analogue of **5** binds an organic-soluble cellobiose derivative in chloroform/methanol mixtures; see G. Lecollinet, A. P. Dominey, T. Velasco, A. P. Davis, *Angew. Chem.* **2002**, *114*, 4267; *Angew. Chem. Int. Ed.* **2002**, *41*, 4093.
[13] MacroModel 9.9 (MMFFs force field, water GB/SA solvation) accessed through the Maestro 9.2 interface.
[14] For further details, see the Supporting Information.
[15] D. H. Leung, R. G. Bergman, K. N. Raymond, *J. Am. Chem. Soc.* **2008**, *130*, 2798.
[16] E. Klein, Y. Ferrand, N. P. Barwell, A. P. Davis, *Angew. Chem.* **2008**, *120*, 2733; *Angew. Chem. Int. Ed.* **2008**, *47*, 2693.
[17] P. Csermely, R. Palotai, R. Nussinov, *Trends Biochem. Sci.* **2010**, *35*, 539; D. Bucher, B. J. Grant, J. A. McCammon, *Biochemistry* **2011**, *50*, 10530.

 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim