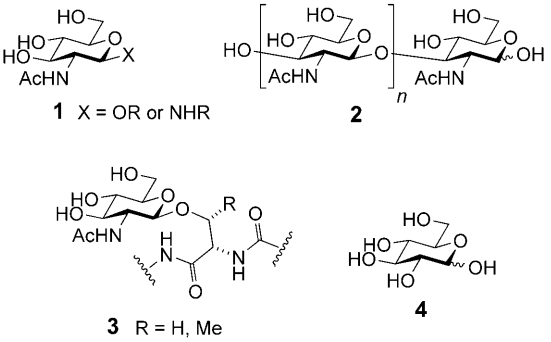
Carbohydrate Recognition

DOI: 10.1002/anie.200804905

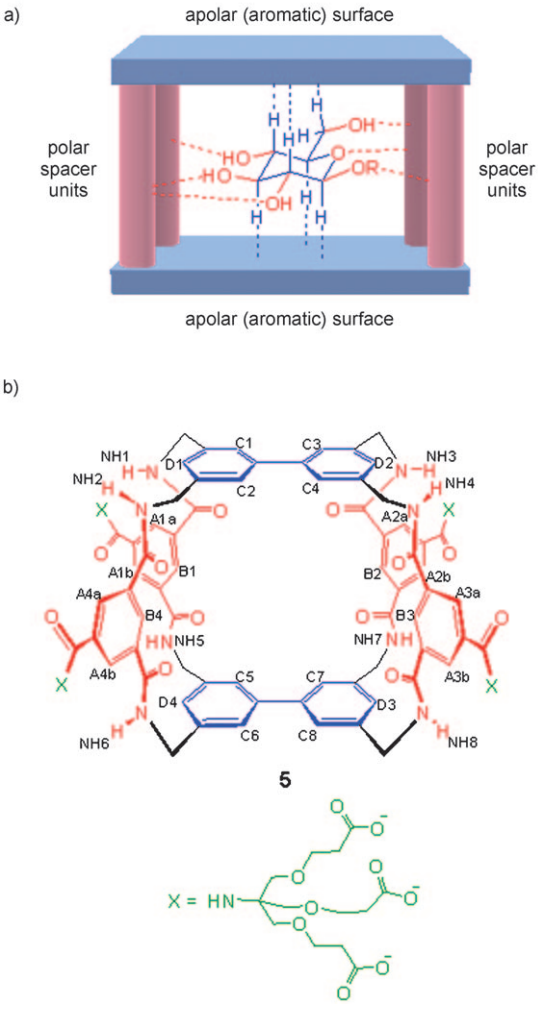
A Synthetic Lectin for O-Linked β-*N*-Acetylglucosamine**

*Yann Ferrand, Emmanuel Klein, Nicholas P. Barwell, Matthew P. Crump, Jesus Jiménez-Barbero, Cristina Vicent, Geert-Jan Boons, Sampat Ingale, and Anthony P. Davis**

β-*N*-Acetyl-D-glucosaminyl (β-GlcNAc, **1**) is a common motif in biological chemistry. It is the monomer building block for chitin (**2**), and occurs frequently in other oligosaccharide structures. It also plays a unique role in protein regulation through linkage to the hydroxy group of serine or threonine as in **3**. This "O–GlcNAc"[1] posttranslational modification is highly dynamic[2] and draws comparisons with protein phosphorylation as a biological control mechanism. It has been implicated in gene transcription, nuclear trafficking, protein translation,[3] signal transduction,[4] the regulation of protein-protein interactions,[1,4] and the sensing of nutritional levels within the cell.[5] Dysregulation of O–GlcNAc contributes to the aetiology of important human diseases, particularly diabetes and neurological disorders.[2]



Research in the Bristol group is aimed at biomimetic carbohydrate receptors[6] that are capable of binding saccharides in water through noncovalent interactions. This goal is challenging as the common carbohydrates are highly hydrophilic and therefore intrinsically difficult to bind from their natural environment. Indeed, lectins[7] (the main class of natural carbohydrate receptors) often show modest affinities, which are typically in the millimolar range for monosaccharides.[8] We have focused especially on binding the β-glucosyl family of saccharides, characterized by "all-equatorial" arrays of polar functional groups. Our approach is illustrated in Scheme 1 a. Equatorial positioning of the polar groups leaves



Scheme 1. Receptors for all-equatorial carbohydrates. a) Schematic representation of the general strategy. b) Monosaccharide receptor **5**, showing the labeling system used for NMR analysis. Apolar surfaces are shown in blue, spacers in red, and water-solubilizing groups in green.

[*] Dr. Y. Ferrand, Dr. E. Klein, N. P. Barwell, Dr. M. P. Crump, 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@bristol.ac.uk
Prof. J. Jiménez-Barbero
Centro de Investigaciones Biológicas, CSIC, Madrid (Spain)
Dr. C. Vicent
Instituto de Química Organica, CSIC, Madrid (Spain)
Prof. G.-J. Boons, S. Ingale
Complex Carbohydrate Research Center
University of Georgia (USA)

[**] This work was supported by the EU (HPRN-CT-2002-00190), the EPSRC (EP/D060192/1), the National Cancer Institute of the NIH (RO1 CA88986), and the Ministry of Science and Innovation of Spain (CTQ2006-10874-C02).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200804905.

hydrophobic patches on either face of the substrate. A complementary cavity may be formed from two parallel aromatic surfaces that are separated by rigid polar spacers. The aromatic surfaces can bind the CH groups through hydrophobic and CH–π interactions,[9] while the spacers can hydrogen bond to the polar substituents. The concept was realized in the form of **5**, with biphenyl "roof" and "floor"[10] and isophthalamide "pillars". Receptor **5** bound glucose (**4**) successfully in water, but with a rather low affinity ($K_a = 9\,M^{-1}$).

While **5** was originally designed to bind β-glucosyl, it was clear that other "all-equatorial" carbohydrates might also be good substrates for this receptor. Given its biological importance, β-GlcNAc seemed an interesting candidate. We report herein that **5** binds certain O-linked β-GlcNAc units with good selectivity and nearly millimolar affinity. The results eclipse the earlier binding data and redefine **5** as a synthetic lectin for O-GlcNAc.

The binding of **5** to *N*-acetylamino carbohydrates was studied in D_2O by using ^{1}H NMR titrations. Initial experiments were performed with *N*-acetylglucosamine (**6**). Instead of the expected signal movements (as observed for glucose), the aromatic signals of **5** were progressively replaced by a new spectrum, which consisted of at least 21 distinguishable resonances (Figure 1). This result implies complex formation with slow equilibration on the NMR timescale. The additional signals are consistent with the loss of symmetry upon binding; once the carbohydrate enters the cavity, all aromatic protons become inequivalent. The binding constant was measurable through the integral ratios of the bound and unbound receptor signals, and was found to be $56\,M^{-1}$.[11] However, GlcNAc (**6**) is a mixture of anomers, α/β = 64:36. On examination of the high-field region, of the spectrum only one bound anomer was detected, this was identified as **6**β from the H1–H2 vicinal coupling constant of 8.8 Hz.[11]

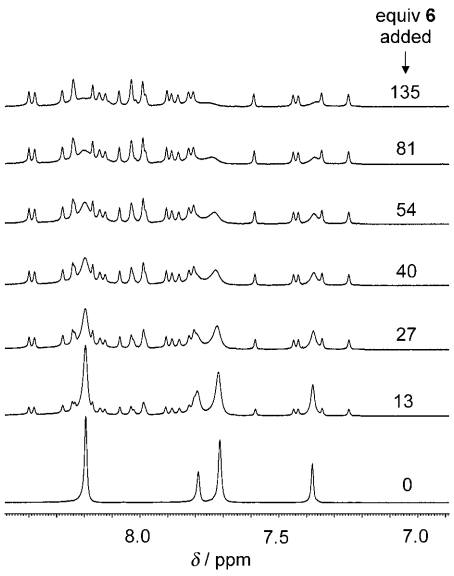
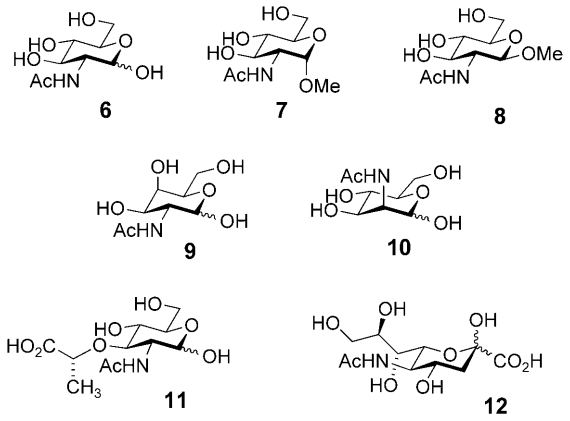


Figure 1. ^{1}H NMR spectra of receptor **5** (0.5 mM) in D_2O after addition of increasing amounts of GlcNAc **6**. Only aromatic signals of the receptor are shown.

Assuming that the new signals observed (Figure 1) belong exclusively to **5**·**6**β, the K_a value for this complex could be recalculated as $156\,M^{-1}$.

Encouraged by this result, we performed ^{1}H NMR titrations for **5** with the anomeric methyl glycosides of GlcNAc (**7** and **8**), *N*-acetylgalactosamine (**9**), *N*-acetylmannosamine (**10**), *N*-acetylmuramic acid (**11**), and *N*-acetylneuraminic (sialic) acid (**12**). Most remarkable were the results for **8**,[11]



the simplest model for the O-GlcNAc protein modification. Again, the spectra contained signals from both receptor and complex, and could be integrated to obtain the association constant K_a. In this case the K_a value was $630\,M^{-1}$, which is nearly two orders of magnitude higher than that reported earlier for glucose and well within the range reported for natural lectins.[8] Confirmation was provided by isothermal titration calorimetry (ITC), which gave $K_a = 635\,M^{-1}$, $\Delta H = -1.92\,kcal\,mol^{-1}$ and $T\Delta S = 1.90\,kcal\,mol^{-1}$. In contrast, **5** showed low affinities for the other new substrates.[11] Binding to **7** could not be quantified by NMR spectroscopy because of signal broadening (this probably reflects an intermediate exchange rate on the NMR timescale). However induced circular dichroism (ICD) titration data fit well to a 1:1 binding model with $K_a = 24\,M^{-1}$. For **9** and **10**, K_a was too low to be measured accurately, but was roughly estimated as $2\,M^{-1}$ in each case. For **11** and **12**, addition of the carbohydrate had no effect on the ^{1}H NMR spectrum of **5**.

The NMR spectra of complex **5**·**8** proved suitable for detailed structural studies.[11] The signals for the bound carbohydrate **8** were readily observed and assigned through 1D spectra as well as 2D ROESY and DQF–COSY. All resonances were shifted upfield relative to the free carbohydrate, which is consistent with the expected chemical shielding effects from the aromatic π electrons. The signals of the CH groups shifted upfield by 1–2.4 ppm upon binding, while the signals of the carbohydrate NH groups shifted upfield by a remarkable 5.1 ppm. For the receptor **5**, the signals of the aromatic CH groups and also the eight signals of the NH groups could be fully assigned by using 2D TOCSY and NOESY. The NOESY information (Figure 2a) was then used to assign 35 intermolecular proton–proton contacts and to estimate approximate distances. Intramolecular NH–CH distances within the receptor were also determined, which

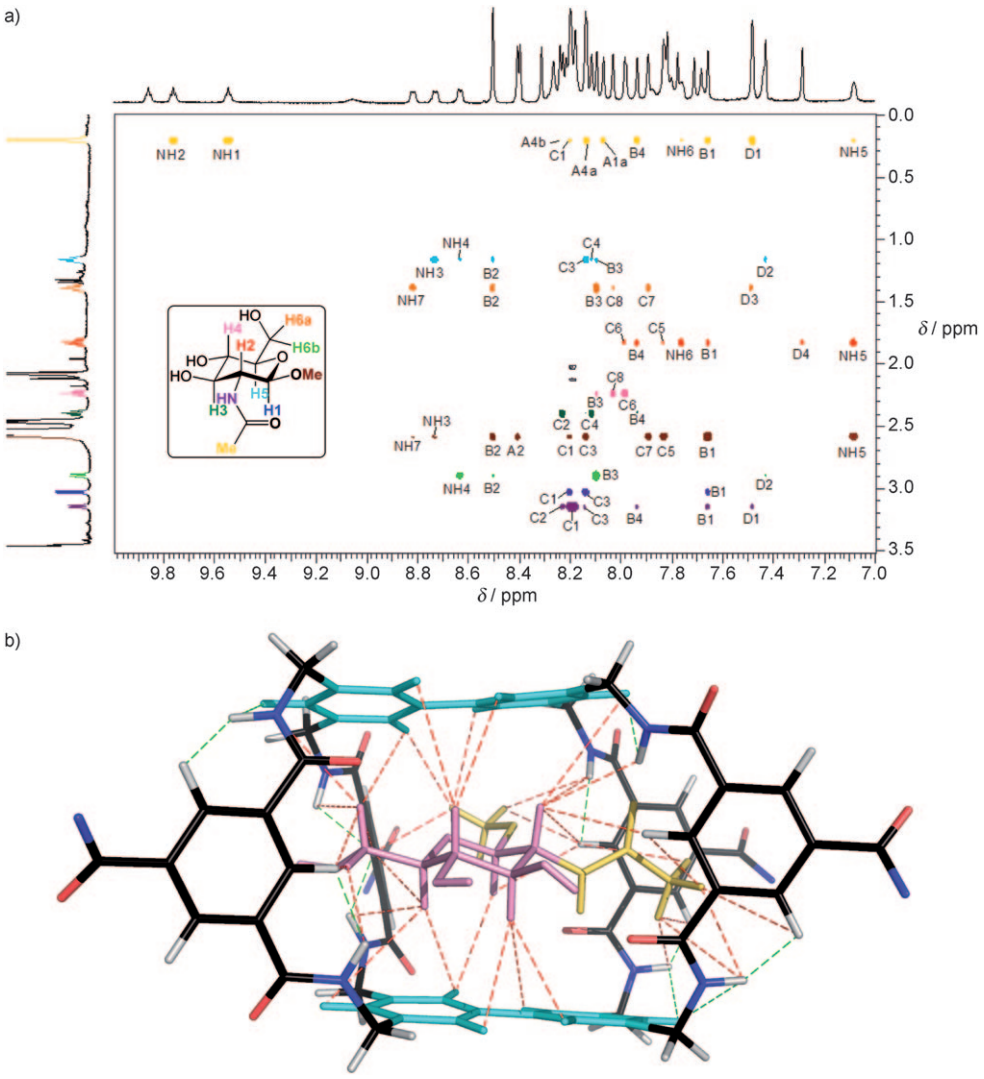


Figure 2. Structure determination of the complex between **5** and **8** by using ^{1}H NMR spectroscopy. a) A portion of the NOESY spectrum of **5** (1 mM) and **8** (8.7 mM) in H_2O/D_2O (9:1) showing intermolecular contacts between carbohydrate and receptor. Mixing time = 150 ms. Each cross-peak is labeled according to the receptor proton (for key, see Scheme 1b), and colored according to the carbohydrate proton (for key, see inset). b) Derived structure for the complex. Isophthalamide spacer atoms are colored according to element (black C, white H, red O, blue N), and the biphenyl units are highlighted in cyan. Carbohydrate **8** is shown as pink, except the OMe and NHAc units, which are highlighted in yellow. Intramolecular and intermolecular nuclear Overhauser effect (nOe) contacts are shown as green and red dotted lines respectively. The water-solubilizing tricarboxylate groups are omitted.

established whether the NH groups pointed into or out of the cavity. The data were incorporated as distance constraints in a Monte Carlo molecular mechanics conformational search. The lowest-energy structure that satisfied all the constraints is shown in Figure 2b. As expected, the carbohydrate is sandwiched between the biphenyl aromatic surfaces. The NHAc group is placed in one of the smaller portals of the cavity, while the methyl group makes good hydrophobic contacts to one biphenyl and two spacer aromatic rings. The carbohydrate CH_2OH group protrudes into the opposite portal, while the methoxy group exits one of the wider openings. Interestingly, the NH atom of the acetamido group is positioned to form an NH–π interaction[12] with a biphenyl aromatic ring, which accounts for its exceptionally low proton chemical shift. There are six conventional intermolecular hydrogen bonds that range from 1.8–2.6 Å in length, two of which involve the oxygen atom of the NHAc group. The NHAc group is thus involved in both polar and apolar interactions, which explains the selectivity for GlcNAc versus glucose. The preference for **8** versus the α anomer **7** or *N*-acetylgalactosamine (**9**) can also be understood. With the NHAc and CH_2OH groups positioned in the narrow portals, the substrate has little room for maneuver. All remaining substituents must be in the equatorial positions to avoid severe steric repulsion between the host and guest.

The binding results for **5** with carbohydrate substrates are shown in Table 1, and reveal a remarkable preference for β-GlcNAc. Selectivity for GlcNAcβ-OMe (**8**) is greater than 100 times that of the affinity of **5** for most common carbohydrates. Even methyl β-D-glucoside and GlcNAcα-OMe **7** are bound with affinities that are more than 20 times lower than that of **8**. The selective recognition of β-GlcNAc has particular importance for the study of the O-GlcNAc protein modification. We were thus interested to see whether **5** would be effective with a more realistic O-GlcNAc model. We therefore prepared the glycopeptide **13**, based on a sequence from casein kinase II (CK II), which is known to be subject to O-GlcNAcylation.[13] Titration of **13** into **5** produced similar ^{1}H NMR spectroscopic changes to those observed when **5** was titrated against either **6** or **8** (Figure 3). In this case, integration of the spectra gave an association constant of $1040\,M^{-1}$.[11] An ICD titration, which gave $K_a = 1100\,M^{-1}$, was carried out to confirm this result. NOESY spectroscopy of the complex showed connections between the receptor aromatic protons and a series of signals from the substrate, and COSY and TOCSY spectra showed that these substrate protons belonged to the GlcNAc residue. It is thus clear that the receptor binds **13** through the carbohydrate unit. Control experiments with aglycosyl pep-

Table 1: Association constants K_a for receptor **5** with carbohydrate substrates in aqueous solution.[a]

Carbohydrate	K_a [M^{-1}]
GlcNAcβ-OMe **8**	630
GlcNAc **6** (α:β = 64:36)	56
methyl β-D-glucoside	28
GlcNAcα-OMe **7**	24[b]
D-cellobiose	17[c]
D-glucose **4**	9[c]
2-deoxy-D-glucose	7[c]
methyl α-D-glucoside	7[c]
D-xylose	5[c]
D-ribose	3[c]
D-galactose	2[c]
L-fucose	2[c]
N-acetyl-D-galactosamine **9**	2[c]
N-acetyl-D-mannosamine **10**	2[c]
D-arabinose	2[c]
D-lyxose	≤2[c]
D-mannose	≤2[c]
L-rhamnose	≤2[c]
D-maltose	≤2[c]
D-lactose	≤2[c]
N-acetyl-D-muramic acid **11**	0
N-acetyl-D-neuraminic acid **12**	0

[a] Measured by ^{1}H NMR titration in D_2O unless otherwise indicated. [b] Measured by induced circular dichroism. [c] Values for these substrates were reported previously; for details see Ref. [10]

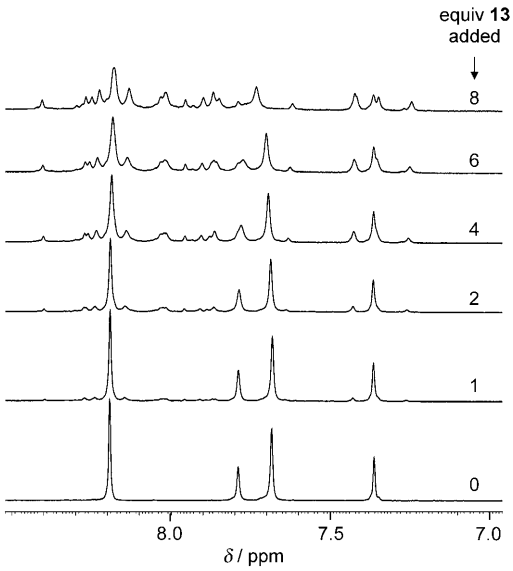
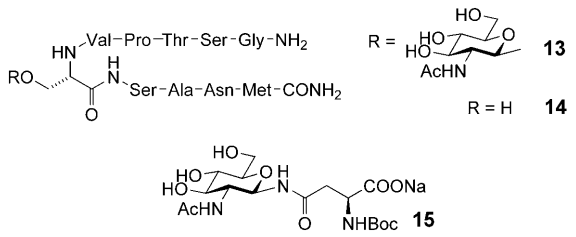


Figure 3. ^{1}H NMR spectra of receptor **5** (0.25 mM) in D_2O after addition of increasing amounts of glycopeptide **13**. Only the aromatic signals of the receptor are shown.

tide **14** supported this conclusion. Addition of **14** to **5** caused no detectable ICD signal, and only small movements of the ^{1}H NMR signals.[14]

As described above, β-GlcNAc may also be attached to itself in chitin **2**, and to asparagine side chains in N-linked glycoproteins. We therefore tested *N*,*N*′-diacetylchitobiose (**2**, $n = 1$) and the β-GlcNAc asparagine derivative **15** as substrates. The former caused no change to the ^{1}H NMR



spectrum of **5**, so appears not to bind at all. We presume that the second (reducing) GlcNAc unit is too bulky to replace the methoxy group in Figure 2b. N-Linked derivative **15** gave broadened spectra, but integration of the signals was possible to give $K_a = 4\ M^{-1}$.[11] This surprisingly low affinity was supported by a competition study in which **8** was shown to displace a 15-fold excess of **15** from the receptor.[11]

It is interesting to compare **5** with the lectins commonly used to bind β-GlcNAc, as lectins are important tools for biology and biomedicine. The best-studied GlcNAc-binding lectin is wheat germ agglutinin (WGA). Some binding constants for WGA are shown in Table 2, in which they are

Table 2: Comparison between the natural lectin Wheat Germ Agglutinin (WGA) and receptor **5**.

Carbohydrate	K_a(WGA) [M^{-1}]	K_a(receptor **5**) [M^{-1}]
GlcNAc **6**	410[a]	56
GlcNAcα-OMe **7**	480[b]	24
GlcNAcβ-OMe **8**	730[b]	630
N-acetyl-D-galactosamine **9**	60[b]	2
N-acetyl-D-mannosamine **10**	60[b]	2
N-acetyl-D-neuraminic acid **12**	560[c]	0
N,*N*′-diacetylchitobiose **2** ($n = 1$)	5300[a]	0

[a] Measured by ITC.[15] [b] Estimated from published data[18] on the inhibition of WGA-induced precipitation, assuming the value shown above for GlcNAc **6**. [c] Measured by ^{1}H NMR titration.[19]

compared with the results for **5**. In terms of affinity for a single β-GlcNAc unit (as represented by **8**), the systems are very similar. However, receptor **5** is far more discriminatory than the protein, **5** shows a higher selectivity for **8** against α-anomer **7** and the other *N*-acetylaminosugars (especially the common oligosaccharide chain terminus **12**). With chitosaccharide **2** ($n = 1$), the contrast between the systems is especially strong, as WGA binds preferentially to the chitosaccharide (and higher oligomers[15]), while **5** shows no affinity at all for these substrates. Other GlcNAc-binding lectins have become available more recently,[16] notably *Griffonia (Bandeiraea) simplicifolia* lectin II (GSL II).[17] Comparisons with **5** are more difficult as studies on these proteins are less detailed, but all bind **2** ($n = 1$) far more strongly than can **5**. GSL II binds α-GlcNAc in preference to β-GlcNAc,[17a] which again contrasts strongly with **5**.

In conclusion, we have shown that **5**, previously seen as a weak receptor for β-glucosyl, is in fact a strong and selective receptor for β-GlcNAc. The affinity of **5** for its new target is remarkable. Although it is matched by our related disacchar-

ide receptor,[6c] the latter benefits from a larger binding surface. The selectivity of **5** is also unusual. Not only does it discriminate between β-GlcNAc and other monosaccharide units, but also between β-GlcNAc in different environments (rejecting, for example, **2** ($n=1$) and **15**). Usefully, it does bind a model of the O-GlcNAc protein modification. It competes well with at least one lectin (WGA) that has been used to bind β-GlcNAc. Although WGA is not the best GlcNAc-binding protein that nature can produce, these are promising results for a fully abiotic synthetic receptor.

Received: October 7, 2008
Published online: December 15, 2008

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

[1] N. E. Zachara, G. W. Hart, *Chem. Rev.* **2002**, *102*, 431.
[2] G. W. Hart, M. P. Housley, C. Slawson, *Nature* **2007**, *446*, 1017.
[3] F. I. Comer, G. W. Hart, *J. Biol. Chem.* **2000**, *275*, 29179.
[4] L. Wells, K. Vosseller, G. W. Hart, *Science* **2001**, *291*, 2376.
[5] N. E. Zachara, G. W. Hart, *Biochim. Biophys. Acta Gen. Subj.* **2004**, *1673*, 13.
[6] For reviews on synthetic carbohydrate receptors, see: 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. For recent developments, see: c) Y. Ferrand, M. P. Crump, A. P. Davis, *Science* **2007**, *318*, 619; d) E. Klein, Y. Ferrand, N. P. Barwell, A. P. Davis, *Angew. Chem.* **2008**, *120*, 2733; *Angew. Chem. Int. Ed.* **2008**, *47*, 2693; e) P. B. Palde, P. C. Gareiss, B. L. Miller, *J. Am. Chem. Soc.* **2008**, *130*, 9566; f) C. He, Z. H. Lin, Z. He, C. Y. Duan, C. H. Xu, Z. M. Wang, C. H. Yan, *Angew. Chem.* **2008**, *120*, 891; *Angew. Chem. Int. Ed.* **2008**, *47*, 877; g) M. Waki, H. Abe, M. Inouye, *Angew. Chem.* **2007**, *119*, 3119; *Angew. Chem. Int. Ed.* **2007**, *46*, 3059; h) T. Reenberg, N. Nyberg, J. O. Duus, J. L. J. van Dongen, M. Meldal, *Eur. J. Org. Chem.* **2007**, 5003; i) C. Nativi, M. Cacciarini, O. Francesconi, A. Vacca, G. Moneti, A. Ienco, S. Roelens, *J. Am. Chem. Soc.* **2007**, *129*, 4377; j) M. Mazik, M. Kuschel, *Chem. Eur. J.* **2008**, *14*, 2405.
[7] 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; N. Sharon, H. Lis, *Glycobiology* **2004**, *14*, 53R; R. M. Ambrosi, N. R. Cameron, B. G. Davis, *Org. Biomol. Chem.* **2005**, *3*, 1593.
[8] E. J. Toone, *Curr. Opin. Struct. Biol.* **1994**, *4*, 719.
[9] M. C. Fernández-Alonso, F. J. Cañada, J. Jiménez-Barbero, G. Cuevas, *J. Am. Chem. Soc.* **2005**, *127*, 7379.
[10] E. Klein, M. P. Crump, A. P. Davis, *Angew. Chem.* **2005**, *117*, 302; *Angew. Chem. Int. Ed.* **2005**, *44*, 298.
[11] For details see the Supporting Information.
[12] M. Levitt, M. F. Perutz, *J. Mol. Biol.* **1988**, *201*, 751; C. Allott, H. Adams, P. L. Bernad, C. A. Hunter, C. Rotger, J. A. Thomas, *Chem. Commun.* **1998**, 2449.
[13] L. K. Kreppel, G. W. Hart, *J. Biol. Chem.* **1999**, *274*, 32015.
[14] The NMR signal movements were almost linearly proportional to the concentration of added **14**, which indicates a very small binding constant.[11]
[15] G. Bains, R. T. Lee, Y. C. Lee, E. Freire, *Biochemistry* **1992**, *31*, 12624.
[16] *Vector Laboratories Catalogue 2007/2008*, Vector Laboratories, Burlingame, CA, **2007**, p. 81.
[17] a) S. Ebisu, P. N. S. Iyer, I. J. Goldstein, *Carbohydr. Res.* **1978**, *61*, 129; b) S. Nakamura-Tsuruta, J. Kominami, M. Kamei, Y. Koyama, T. Suzuki, M. Isemura, J. Hirabayashi, *J. Biochem.* **2006**, *140*, 285.
[18] A. K. Allen, A. Neuberger, N. Sharon, *Biochem. J.* **1973**, *131*, 155; I. J. Goldstein, S. Hammarstrom, G. Sundblad, *Biochim. Biophys. Acta Protein Struct.* **1975**, *405*, 53.
[19] F. Jordan, E. Bassett, W. R. Redwood, *Biochem. Biophys. Res. Commun.* **1977**, *75*, 1015.

