Published in final edited form as:

Angew Chem Int Ed Engl. 2009 ; 48(10): 1775–1779. 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

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 9298611, Anthony.Davis@bristol.ac.uk

Prof. J. Jiménez-Barbero, Centro de Investigaciones Biológicas, CSIC, c/ Ramiro de Maeztu 9, Madrid 28040 (Spain)

Dr. C. Vicent, Instituto de Química Organica, CSIC, c/ Juan de la Cierva 3, Madrid 28006 (Spain)

Prof. G.-J. Boons, S. Ingale, Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Road, Athens, GA 30602 (USA)

Keywords

Molecular recognition; supramolecular chemistry; receptors; carbohydrates; biomimetic

β -N-Acetyl-D-glucosamine (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 hydroxyl of serine or threonine as in **3**. This “O-GlcNAc”[1] post-translational 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] capable of binding saccharides in water through non-covalent interactions. This goal is challenging, as the common carbohydrates are highly hydrophilic and therefore intrinsically difficult to bind from their natural environment. Indeed, natural carbohydrate receptors (lectins)[7] often

Correspondence to: Anthony P. Davis.

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

show modest affinities, typically millimolar for monosaccharides.[8] We have focused especially on binding the β -glucosyl family of saccharides, characterised by all-equatorial arrays of polar functional groups. Our approach is illustrated in Figure 1a. Equatorial positioning of the polar groups leaves hydrophobic patches on either face of the substrate. A complementary cavity may be formed from two parallel aromatic surfaces, 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 realised in the form of **5**, with biphenyl “roof” and “floor”[10] and isophthalamide “pillars”. Receptor **5** bound glucose (**4**) successfully in water, but with rather low affinity ($K_a = 9 \text{ M}^{-1}$).

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

The binding of **5** to *N*-acetylamino carbohydrates was studied in D_2O by ^1H NMR titrations. Initial experiments were performed with *N*-acetylglucosamine (**6**). Instead of the expected signal movements (as observed for glucose), the aromatic signals due to **5** were replaced progressively by a new spectrum consisting of at least 21 distinguishable resonances (Figure 2). This result implies complex formation with slow equilibration on the NMR timescale. The additional signals are consistent with the loss of symmetry on binding; once the carbohydrate enters the cavity, all aromatic protons become inequivalent. The binding constant was measurable through integration of the signals for bound vs. unbound receptor, and was found to be 56 M^{-1} . [11] However, GlcNAc **6** is a mixture of anomers, $\alpha:\beta = 64:36$. On examination of the high field region of the spectrum only one bound anomer was detected, identified as **6 β** from the H1-H2 vicinal coupling constant of 8.8 Hz.[11] Assuming that the new signals observed (Fig. 2) 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 ^1H NMR titrations for **5** + the anomeric methyl glycosides of GlcNAc (α **7**, β **8**), *N*-acetylglactosamine (**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 value was 630 M^{-1} , 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 $K_a = 635 \text{ M}^{-1}$, $\Delta H = -1.92 \text{ kcal mol}^{-1}$ and $T\Delta S = 1.90 \text{ kcal mol}^{-1}$. In contrast,

5 showed low affinities for the other new substrates.[11] Binding to **7** could not be quantified by NMR due to signal broadening (probably reflecting an intermediate exchange rate on the NMR timescale). However induced circular dichroism (ICD) titration data fit well to a 1:1 binding model with $K_a = 24 \text{ M}^{-1}$. 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 ^1H 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, and 2D ROESY and DQF-COSY. All resonances were shifted upfield relative to the free carbohydrate, consistent with the expected chemical shielding effects from the aromatic π -electrons. The *CH* signals were moved by 1 – 2.4 p.p.m. on binding, while the carbohydrate *NH* shifted upfield by a remarkable 5.1 p.p.m. For the receptor **5**, the aromatic *CH* and also the 8 *NH* signals could be fully assigned using 2D TOCSY and NOESY. The NOESY information (Fig. 3a) 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, establishing whether *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 which satisfied all the constraints is shown in Figure 3b. 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, with the methyl group making good hydrophobic contacts to one biphenyl and two spacer aromatic rings. The carbohydrate CH_2OH protrudes into the opposite portal, while the methoxy group exits one of the wider openings. Interestingly, the acetamido *NH* is positioned to form an *NH- π* interaction[12] with a biphenyl aromatic ring, accounting for its exceptionally low proton chemical shift. There are 6 conventional intermolecular H-bonds ranging from 1.8–2.6 Å in length, two involving the *NHAc* oxygen. The *NHAc* group is thus involved in both polar and apolar interactions, explaining the selectivity for GlcNAc vs. glucose. The preference for **8** vs. 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 manoeuvre. All remaining substituents must be equatorial to avoid severe steric repulsion between host and guest.

The binding results for **5** + carbohydrate substrates are collected in Table 1, and reveal a remarkable preference for β -GlcNAc. Selectivity for **8** vs. most common carbohydrates is $\geq 100:1$. Even methyl β -D-glucoside is bound with >20 times lower affinity, as is GlcNAc- α -OMe **7**. The selective recognition of β -GlcNAc has particular importance for the study of the O-GlcNAc protein modification, so we were interested to find 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 ^1H NMR spectroscopic changes to those observed for **6** and **8** (Fig. 4). In this case, integration of the spectra gave an association constant of 1040 M^{-1} . [11] A confirmatory study employing ICD titration gave $K_a = 1100 \text{ M}^{-1}$. 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** via the carbohydrate unit. Control experiments with aglycosyl peptide **14** supported this conclusion. Addition of **14** to **5** caused no detectable ICD signal, and just small movements of ^1H NMR signals.[14]

As mentioned earlier, 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 ^1H 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 Fig. 3b. N-Linked derivative **15** gave broadened spectra, but integration was possible to give $K_a = 4 \text{ M}^{-1}$.^[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]

As described above, lectins are important tools for biology and biomedicine. It is interesting to compare **5** with the lectins commonly used to bind β -GlcNAc. The best-studied is Wheat Germ Agglutinin (WGA). Some binding constants for WGA are shown in Table 2, where they are compared to 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, selecting more effectively 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. WGA prefers the chitosaccharide (and higher oligomers^[15]), while **5** shows no affinity at all. 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 does **5**. GSL II prefers α -GlcNAc to β -GlcNAc,^[16a] again contrasting 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 disaccharide 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) which 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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).

References

1. Zachara NE, Hart GW. *Chem Rev* 2002;102:431. [PubMed: 11841249]
2. Hart GW, Housley MP, Slawson C. *Nature* 2007;446:1017. [PubMed: 17460662]
3. Comer FI, Hart GW. *J Biol Chem* 2000;275:29179. [PubMed: 10924527]
4. Wells L, Vosseller K, Hart GW. *Science* 2001;291:2376. [PubMed: 11269319]
5. Zachara NE, Hart GW. *Biochimica Et Biophysica Acta-General Subjects* 2004;1673:13.
6. Reviews on synthetic carbohydrate receptors: a) Davis AP, Wareham RS. *Angew Chem* 1999;111:3160. *Angew Chem Int Ed* 1999;38:2978. b) Davis, AP.; James, TD. *Functional Synthetic Receptors*. Schrader, T.; Hamilton, AD., editors. Vol. 45. Wiley-VCH; Weinheim: 2005. Recent developments: c) Ferrand Y, Crump MP, Davis AP. *Science* 2007;318:619. [PubMed: 17962557] d) Klein E, Ferrand Y, Barwell NP, Davis AP. *Angew Chem* 2008;120:2733. *Angew Chem Int Ed* 2008;48:2693. e) Palde PB, Gareiss PC, Miller BL. *J Am Chem Soc* 2008;130:9566. [PubMed: 18576640] f) He C, Lin ZH, He Z, Duan CY, Xu CH, Wang ZM, Yan CH. *Angew Chem* 2008;120:891. *Angew Chem Int Ed* 2008;47:877. g) Waki M, Abe H, Inouye M. *Angew Chem* 2007;119:3119. *Angew Chem Int Ed* 2007;46:3059. h) Reenberg T, Nyberg N, Duus JO, van Dongen JIJ, Meldal M. *Eur J Org Chem* 2007:5003. i) Nativi C, Cacciarini M, Francesconi O, Vacca A, Moneti G, Ienco A, Roelens S. *J Am Chem Soc* 2007;129:4377. [PubMed: 17362009] j) Mazik M, Kuschel M. *Chem Eur J* 2008;14:2405.
7. Gabius HJ, Siebert HC, Andre S, Jimenez-Barbero J, Rudiger H. *Chem BioChem* 2004;5:740. Dwek RA, Butters TD. *Chem Rev* 2002;102:283. Bertozzi CR, Kiessling LL. *Science* 2001;291:2357. [PubMed: 11269316] Sharon N, Lis H. *Glycobiology* 2004;14:53. Ambrosi RM, Cameron NR, Davis BG. *Org Biomol Chem* 2005;3:1593. [PubMed: 15858635]
8. Toone EJ. *Curr Opin Struct Biol* 1994;4:719.
9. Fernández-Alonso MC, Cañada FJ, Jiménez-Barbero J, Cuevas G. *J Am Chem Soc* 2005;127:7379. [PubMed: 15898786]
10. Klein E, Crump MP, Davis AP. *Angew Chem* 2005;117:302. *Angew Chem Int Ed* 2005;44:298.
11. For details see Supporting Information.
12. Levitt M, Perutz MF. *J Mol Biol* 1988;201:751. [PubMed: 3172202] Allott C, Adams H, Bernad PL, Hunter CA, Rotger C, Thomas JA. *Chem Commun* 1998:2449.
13. Kreppel LK, Hart GW. *J Biol Chem* 1999;274:32015. [PubMed: 10542233]
14. The NMR signal movements were almost linear with [**14**], indicating a very small binding constant.[11]
15. Bains G, Lee RT, Lee YC, Freire E. *Biochemistry* 1992;31:12624. [PubMed: 1472499]
16. Vector Laboratories Catalogue, 2007/2008. Vector Laboratories; Burlingame, CA: 2007. p. 81
17. a) Ebisu S, Iyer PNS, Goldstein IJ. *Carbohydrate Research* 1978;61:129. b) Nakamura-Tsuruta S, Kominami J, Kamei M, Koyama Y, Suzuki T, Isemura M, Hirabayashi J. *Journal of Biochemistry* 2006;140:285. [PubMed: 16835257]
18. Allen AK, Neuberger A, Sharon N. *Biochemical Journal* 1973;131:155. [PubMed: 4737292] Goldstein IJ, Hammarstrom S, Sundblad G. *Biochimica Et Biophysica Acta* 1975;405:53. [PubMed: 1174568]
19. Jordan F, Bassett E, Redwood WR. *Biochem and Biophys Res Comm* 1977;75:1015. [PubMed: 871315]

Figure 1. Receptors for all-equatorial carbohydrates. a) Cartoon depicting the general strategy. b) Monosaccharide receptor **5**, showing the labelling system used for NMR analysis. Apolar surfaces are shown as blue, spacers red and water-solubilising groups as green.

Figure 2. ^1H NMR spectra for addition of GlcNAc **6** to receptor **5** (0.5 mM) in D_2O , observing the signals due to aromatic protons in the receptor.

Figure 3.

^1H NMR structure determination of the complex between **5** and **8**. a) A portion of the NOESY spectrum of **5** (1 mM) + **8** (8.7 mM) in $\text{H}_2\text{O}/\text{D}_2\text{O}$ (9:1) showing intermolecular contacts between carbohydrate and receptor. Mixing time = 150 ms. Each cross peak is labelled according to the receptor proton (for key, see Fig. 1b), and coloured according to the carbohydrate proton (for key, see inset). b) Derived structure for the complex. Isophthalamide spacer atoms are coloured according to element (C = black, H = white, O = red, N = blue), 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 nOe contacts are shown as green and red dotted lines respectively. The water-solubilising tricarboxylate groups are omitted.

Figure 4. ^1H NMR spectra for addition of glycopeptide **13** to **5** (0.25 mM) in D_2O , observing the signals due to aromatic protons in the receptor.

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

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

^a Measured by ¹H NMR titration in D₂O unless otherwise indicated. Values for italicised substrates were reported previously; for details see ref. [10].

^b Measured by Induced Circular Dichroism.

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

Carbohydrate	WGA K_a (M^{-1})	Receptor $5K_a$ (M^{-1})
GlcNAc 6	410 ^a	56
GlcNAc α -OMe 7	480 ^b	24
GlcNAc β -OMe 8	730 ^{b†}	630
<i>N</i> -acetyl-D-galactosamine 9	60 ^{b†}	2
<i>N</i> -acetyl-D-mannosamine 10	60 ^{b†}	2
<i>N</i> -acetyl-D-neuraminic acid 12	560 ^c	0
<i>N,N'</i> -diacetylchitobiose 2 (n = 1)	5300 ^a	0

^a Measured by ITC.[14]^b Estimated from published data[18] on the inhibition of WGA-induced precipitation, assuming the value shown above for GlcNAc **6**.^c Measured by ¹H NMR titration.[19]