

A Simple Biomimetic Receptor Selectively Recognizing the GlcNAc₂ Disaccharide in Water

Oscar Francesconi,* Francesco Milanesi, Cristina Nativi, and Stefano Roelens*

In memory of Professor Alberto Vacca

Abstract: GlcNAc₂ is the core disaccharide fragment present in *N*-glycans exposed on the surface of enveloped viruses of high health concern, such as coronaviruses. Because *N*-glycans are directly involved in the docking of viruses to host cells, recognition of GlcNAc₂ by a biomimetic receptor may be a convenient alternative to the use of lectins to interfere with viral entry and infection. Herein, we describe a simple biomimetic receptor recognizing the methyl- β -glycoside of GlcNAc₂ in water with an unprecedented affinity of 160 μ M, exceeding that of more structurally complex receptors reported in the literature. The tweezers-shaped acyclic structure exhibits marked selectivity among structurally related disaccharides, and complete discrimination between mono- and disaccharides. Molecular modelling calculations supported by NOE data provided a three-dimensional description of the binding mode, shedding light on the origin of the affinities and selectivities exhibited by the receptor.

Enveloped viruses are a broad class of highly glycosylated viroids of high health concern, including coronaviruses (SARS-CoV-2, SARS-CoV and MERS), retroviruses (HIV and hepatitis B), orthomyxoviruses (influenza A-C), flaviviruses (dengue, hepatitis C, yellow fever, Zika) and filoviruses (Ebola and Marburg fever), among others.^[1] Viral adhesion to host cells is often mediated by specific carbohydrate-protein interactions, which occur through the glycans exposed on the surface of the viral envelope.^[2] Biomimetic receptors for carbohydrates targeting these saccharides may inhibit virus-cell interaction, thereby preventing viral entry and infection.^[3,4] In this context, among biologically relevant oligosaccharides, *N,N'*-diacetylchitobiose (GlcNAc₂) holds a pivotal

role, because is part of the highly conserved GlcNAc₂Man₃ core fragment of *N*-glycans present on the surface of enveloped viruses, constituting the disaccharidic unit N-linked to membrane proteins through an asparagine residue, which often get exposed by mannoside deletions due to virus mutations.^[5,6] Unsurprisingly, GlcNAc-binding lectins, such as NICTABA from *Nicotiana tabacum* and *Urtica dioica* Agglutinin (UDA), which target GlcNAc₂ at the stem of *N*-glycosylation sites, possess a broad-spectrum activity against several families of enveloped viruses.^[7] Thus, effective molecular recognition of GlcNAc₂ in water by a simple and easily accessible biomimetic receptor can potentially represent a convenient alternative to natural lectins, because of advantages in terms of availability, molecular weight, purity, stability and immunogenicity.

Selective recognition of neutral glycans by biomimetic receptors in physiological medium represents a major challenge of current research,^[4,8] because water is a strong competitor for recognition of polar molecules such as carbohydrates.^[9] Nevertheless, in the last few years, significant steps forward have been made in the design of receptors for mono- and oligosaccharides, mainly developing appropriately sized macrocyclic architectures.^[8,10] The main drawback of the latter strategy is that macrocyclic architectures must be precisely tailored on specific saccharidic targets and often require lengthy multistep syntheses with relatively low overall yields, due to the critical macrocyclization step.

We have recently developed a new generation biomimetic receptor for monosaccharides in water, by assembling into a macrocyclic architecture a tridentate hydrogen binding motif (1,8-diaminocarbazole endowed with two hydrosolubilizing phosphonate groups) with anthracene moieties providing extended CH- π interactions with the saccharidic backbone.^[11,12] Interestingly, the corresponding adaptive tweezers-shaped liposoluble receptor proved to effectively recognize biologically relevant xanthines in organic solvents.^[13] Following the idea that an acyclic adaptive structure may accommodate disaccharides more effectively than its macrocyclic counterpart, we tested receptor **1**, the hydrosoluble version of the parent receptor **2**,^[13] vs. a set of mono- and disaccharides in water, in the belief that effective recognition of disaccharides may be achieved by a simple, easily available structure.^[14] We report here that this is true indeed, and that **1** is not only a simple and easily accessible receptor for disaccharides, but also the most effective biomimetic receptor for GlcNAc₂ in the literature up to date.

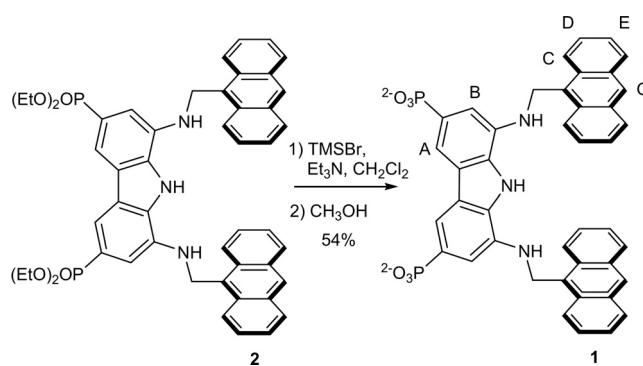
Compound **1** was easily obtained by hydrolysis of the previously reported ester **2** (Scheme 1).^[13] **1** is freely soluble in

[*] Dr. O. Francesconi, F. Milanesi, Prof. C. Nativi, Dr. S. Roelens
 Department of Chemistry “Ugo Schiff” and INSTM, University of Florence, Polo Scientifico e Tecnologico 50019 Sesto Fiorentino, Firenze (Italy)
 E-mail: oscar.francesconi@unifi.it
 stefano.roelens@unifi.it

F. Milanesi
 Magnetic Resonance Center CERM
 Via L. Sacconi 6, 50019 Sesto Fiorentino, Firenze (Italy)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
<https://doi.org/10.1002/anie.202100560>.

© 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

**Scheme 1.** Synthesis of receptor **1** and proton labels.

water under both, mild alkaline (pH 11) and physiological (pH 7.4) conditions, whereas precipitates at acidic pH, due to high degree of protonation of phosphonate groups. Receptor **1** shows sharp ^1H NMR signals at low concentration values ($5 \times 10^4 \text{ M}$), broadening at higher values at pH 7.4, but not at pH 11, suggesting concentration-dependent self-association, supported by chemical shift changes.

The binding properties of receptor **1** were qualitatively screened by ^1H NMR spectroscopy toward a set of monosaccharides, including glucose, rhamnose, fucose, xylose, sialic acid, α and β methyl glucosides, galactosides, mannosides, and *N*-acetylglucosamine (Figure 1a), together with a set of

disaccharides, including sucrose (Suc), trehalose (Tre), cellobiose (CeB), maltose (Mal) and lactose (Lac) (Figure 1b) by monitoring the shifts of the proton signals of the sugar upon addition of an equimolar amount of **1**. Surprisingly, little ($\Delta\delta \leq 0.03 \text{ ppm}$) or no variations were observed for all the investigated monosaccharides and for Suc and Tre, whereas a marked upfield shift was observed for CeB, Mal and Lac, which were larger for the β than for the α anomer (Figures S3–S5), reasonably due to the shielding effect of the anthracene moieties in the binding cavity. A concomitant broadening of signals, larger for the CeB, indicated slow chemical exchange, most likely due to strong binding, suggesting a preference for all-equatorial disaccharides.

A quantitative investigation was then carried out by NMR spectroscopy, extending the study to the all-equatorial GlcNAc₂. Because in *N*-glycans the GlcNAc₂ glycoside unit is present as the β anomer, methyl- β -glycosides of GlcNAc₂ (Me β GlcNAc₂), cellobiose (Me β CeB), maltose (Me β Mal) and lactose (Me β Lac) (Figure 1a) were employed, to avoid interconversion equilibria between anomers. Dilution experiments of receptor **1** were preliminary carried out at pH 7.4, fitting a self-association model featuring two clusters, in which the dimer was the predominant species at low concentrations. The fit gave a dimerization constant of $\log\beta_{\text{dim}} = 2.65 \pm 0.07$, most likely due to π stacking of the aromatic moieties, which was set invariant in the nonlinear regression analysis of the glycoside binding experiments. The cumulative association

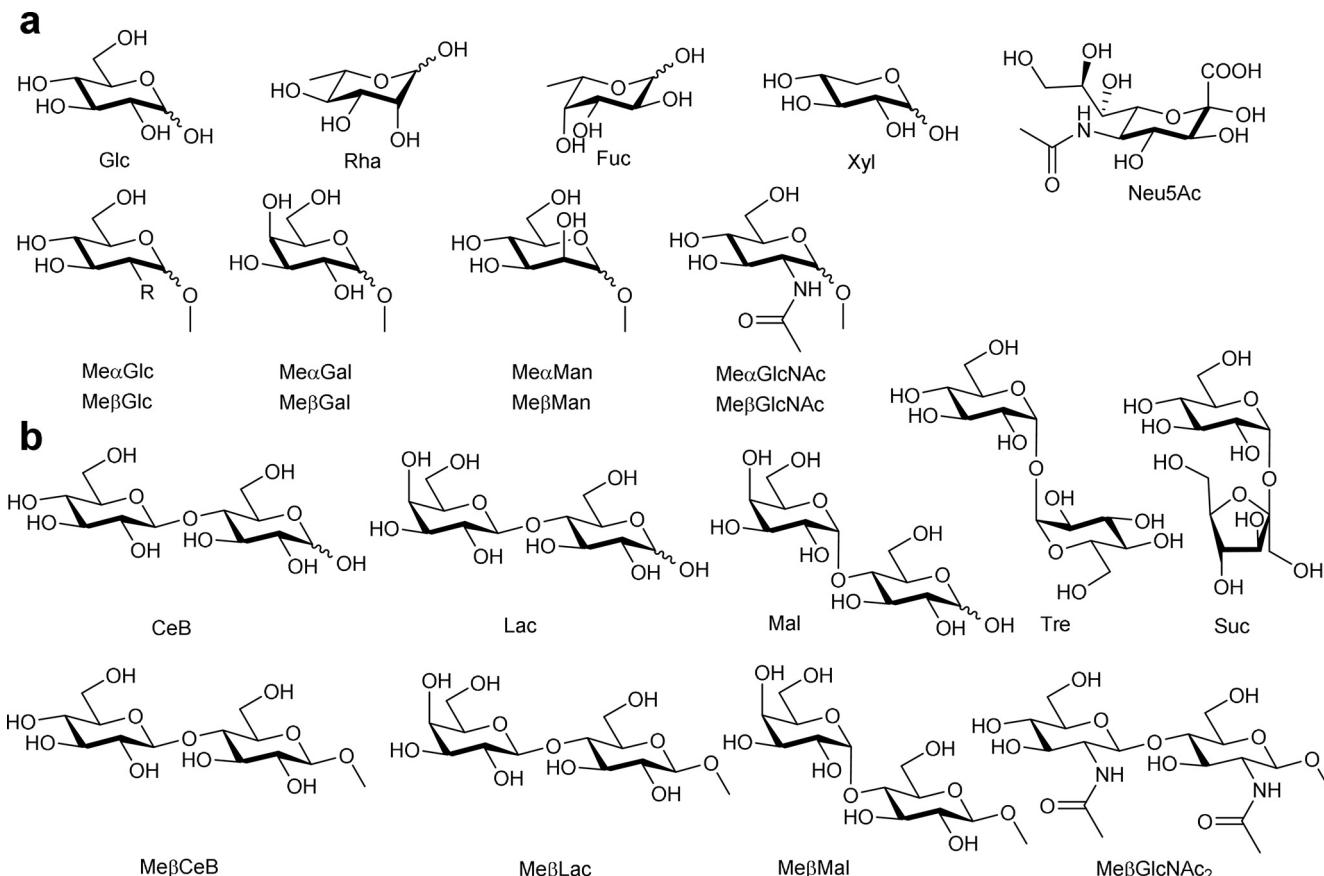
**Figure 1.** Structure of the investigated a) monosaccharides and b) disaccharides and their abbreviations.

Table 1: Cumulative formation constants ($\log \beta_n$)^[a] and intrinsic median binding concentration (BC_{50}^0 , mM)^[b] for receptor to glycoside (R:G) complexes of **1** with methyl glycosides, measured at 298 K from NMR data in D_2O at pD 7.4 and from ITC data in H_2O at pH 7.4.^[c]

Glycoside	R:G	NMR		ITC	
		$\log \beta$	BC_{50}^0	$\log \beta$	BC_{50}^0
Me β CeB	1:1	2.53 ± 0.07	0.94 ± 0.10	2.58 ± 0.03	3.52 ± 2.35
	2:1	6.33 ± 0.06		5.28 ± 0.29	
Me β Mal	1:1	2.27 ± 0.01	31.0 ± 4.4	2.24 ± 0.01	34.4 ± 4.9
	1:1	2.27 ± 0.02	30.8 ± 4.7	2.31 ± 0.02	26.1 ± 4.1
Me β Lac	1:1	3.55 ± 0.04	0.16 ± 0.01	3.49 ± 0.07	0.12 ± 0.03
	2:1	7.35 ± 0.09		7.71 ± 0.22	

[a] Formation constants were obtained by nonlinear least-square regression analysis of NMR and ITC data. [b] Calculated from the $\log \beta$ values using the "BC50 Calculator" program.^[15] [c] Receptor **1** dimerization constant ($\log \beta_{\text{dim}} = 2.65 \pm 0.07$) was set invariant in the nonlinear regression analysis of NMR and ITC data.

constants, reported in Table 1, were measured by 1H NMR titrations in D_2O (pD 7.4) at 298 K, simultaneously fitting the complexation induced shifts of all the available signals to the appropriate association model by non-linear regression analysis. Because multiple binding constants were measured, affinities were assessed through the intrinsic median binding concentration parameter BC_{50}^0 ,^[15] calculated from the measured binding constants and reported in Table 1. Amazingly, results show that receptor **1** binds Me β GlcNAc₂ with an affinity of 160 μM which is unprecedented in the literature for a synthetic receptor. Indeed, to the best of our knowledge, the highest affinity reported to date for Me β GlcNAc₂ by a biomimetic receptor is that observed by Davis and co-workers with a bicyclic polyamidic receptor, showing a 3-fold lower affinity than **1** ($BC_{50}^0 = 455 \mu\text{M}$, $K_a = 2200 \text{ M}^{-1}$).^[16] The affinity of **1** for Me β GlcNAc₂ exceeds that of some lectin-like proteins, such as hevein from *Hevea brasiliensis*, which shows for GlcNAc₂ an affinity one order of magnitude smaller ($BC_{50}^0 = 1.61 \text{ mM}$, $K_a = 620 \pm 50$).^[17] Moreover, receptor **1** exhibits a marked selectivity, showing an affinity for Me β CeB nearly one order of magnitude smaller, and a 200-fold drop of affinity for Me β Mal and Me β Lac. 1H NMR titrations with Me β CeB were also duplicated at pD 11 (Table S1, Supporting Information) and fitted to the association model obtained at pD 7.4. The closely comparable affinities confirmed that the degree of protonation of the phosphonate groups does not affect the binding ability of receptor **1**. Most remarkably, recognition of monosaccharides appears completely depleted, as appreciated from the preliminary screening and from the titration of Me β GlcNAc, in which no significant variation of chemical shifts was observed (Figure S13). Somewhat counterintuitively, a flexible acyclic structure exhibits excellent affinities and selectivities, overriding those of more structurally complicated macrocyclic architectures.^[8]

The binding affinities obtained by NMR were further confirmed by ITC in H_2O at physiological pH. Data from two to three independent titrations run at different reactant concentrations were combined and simultaneously fitted to remove ambiguities in the definition of the binding model. The dimerization constant obtained by NMR dilution experi-

ments at pD 7.4 was set invariant in the nonlinear regression analysis of ITC data. Cumulative binding constants, together with affinity values, were reported in Table 1 for a direct comparison with NMR results. The good agreement between the two independent techniques confirmed the observed affinities. Unfortunately, because of the strong self-association of receptor **1**, ITC measurements did not provide reliable thermodynamic parameters.

To shed light on the origin of the affinities and selectivities exhibited by receptor **1**, a three-dimensional description of the binding mode was attempted for the 1:1 complex of **1** with Me β GlcNAc₂, using molecular modelling calculations supported by NOE data from NMR spectroscopy. NOESY spectra carried out on an equimolar mixture of **1** and Me β GlcNAc₂ showed unambiguous intermolecular NOE contacts between the H-2 and H'-6' protons, belonging to the two monosaccharide units, and the H-C protons of the anthracenes (Figure S20). In addition, NOE contacts were found between the methyl protons of the *N*-acetyl group of the methylglycosidic unit with the H-C protons of the anthracene and the H-A and H-B protons of the carbazole. In contrast, no NOE contacts could be found for the second *N*-acetyl group.

A conformational search carried out on the 1:1 complex returned a single family of minimum energy conformers within 10.0 kJ mol^{-1} from the global minimum. The minimum energy structure depicted in Figure 2 shows the disaccharide entirely located inside the binding cleft, within the two anthracene faces, with the H-2 and H'-6' protons pointing toward the H-C protons of the anthracenes, in agreement with the proximities inferred by NOE contacts (Figure 2a). In addition, the *N*-acetyl group of the methylglycosidic unit faces the diaminocarbazole moiety, pointing the methyl protons toward the H-A and H-B protons, and to one of the H-C protons of the anthracene. From the above model, all O···H interatomic distances shorter than the sum of the van der Waals radii and compliant with hydrogen bonding criteria were calculated and four hydrogen bonding interactions were found between **1** and Me β GlcNAc₂ (Figure 2b): the carbazole NH and one of the aminic NH behave as donors toward OH-3 of the methylglycosidic unit, whereas the other aminic NH of the receptor acts as both, donor to the OH-3 and acceptor from the amidic NH of the *N*-acetyl group. In addition to CH- π interactions with the anthracene units,^[18] the *N*-acetyl group contributes to stabilize the complex through CH- π interactions with the carbazole unit. Most likely, the substantial participation to binding of the *N*-acetyl group may account for the observed selectivity over other disaccharides.

In this context, it is worth mentioning that, interestingly, the binding mode of **1** with Me β GlcNAc₂ is reminiscent of that between hevein and the corresponding trisaccharide chitotriose, as reported by the group of J. Jiménez-Barbero (Figure 3).^[17,19] A close similarity can be appreciated between the binding mode of chitotriose to Tyr30 of hevein and of chitobiose to the aminocarbazole of **1**. Indeed, both receptors engage hydrogen-bonding with the saccharidic OH-3 and with the acetamidic NH, while in both cases the methyl group stabilizes the complex through CH- π interactions with the aromatic ring.

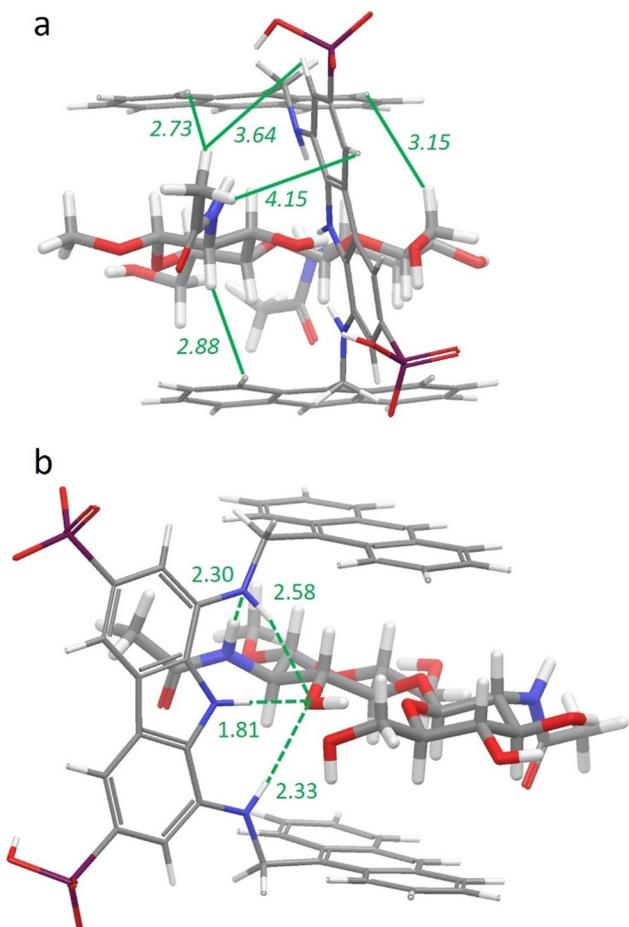


Figure 2. Global minimum structure of the **1** Me β GlcNAc₂ complex in two different projections. a) The strongest intermolecular NOEs found between **1** and Me β GlcNAc₂ are indicated as solid lines with corresponding distances [Å] calculated from the lowest energy conformer. b) Intermolecular hydrogen-bonding interactions found in the calculated structure are indicated as dashed lines with corresponding oxygen/hydrogen and nitrogen/hydrogen distances [Å].

Altogether, the present work shows that molecular recognition of a disaccharide can effectively and selectively be achieved with a well-designed acyclic host featuring an adaptive architecture. The tweezers-shaped receptor **1** fully discriminates disaccharides from monosaccharides, selectively recognizes all-equatorial from non all-equatorial disaccharides, and shows an unprecedented affinity for GlcNAc₂, the core glycosidic fragment of viral *N*-glycans. The hydrogen-bonding and CH– π interactions established by receptor **1** with the *N*-acetyl group most likely account for the selectivity observed for Me β GlcNAc₂ over other all-equatorial disaccharides. Because of simple structure, easy synthetic availability, and accessible structure modifications, receptor **1** stands as a promising tool for saccharide recognition.

Acknowledgements

We are grateful to Prof. J. Jiménez-Barbero for providing the coordinates of the hevein complex. We thank MIUR-Italy

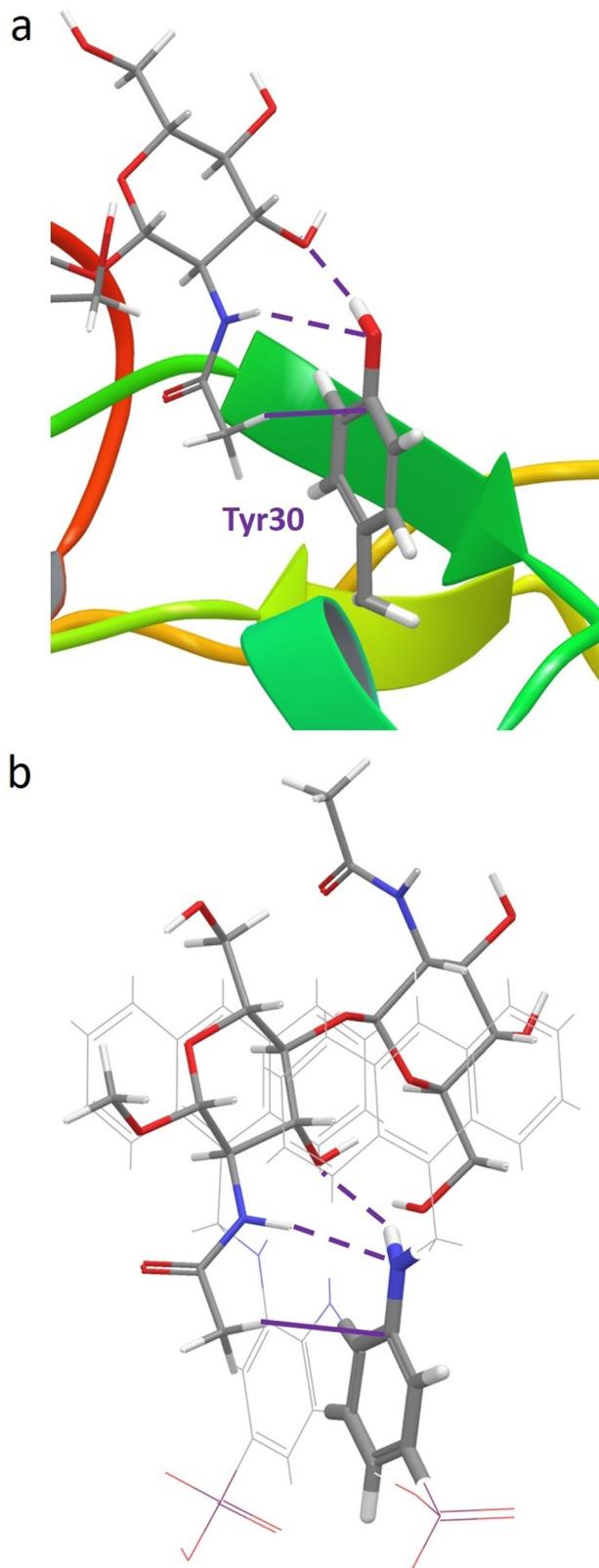


Figure 3. Comparison of the global minimum structures of a) hevein-chitotriose complex and b) **1*** Me β GlcNAc₂ complex. Intermolecular hydrogen-bonding and CH– π interactions involving Tyr30 and the arylamine moiety of **1** are indicated as dashed and solid lines, respectively. Protein backbone was drawn as a ribbon for clarity.

(“Progetto Dipartimenti di Eccellenza 2018–2022” allocated to Department of Chemistry “Ugo Schiff”) and Consorzio Interuniversitario Nazionale di ricerca in Metodologie e Processi Innovativi di Sintesi (CINMPIS). Ente Cassa di Risparmio di Firenze (Italy) is acknowledged for granting an ITC nanocalorimeter and a high-field NMR spectrometer.

Conflict of interest

The authors declare no conflict of interest.

Keywords: carbohydrates · chitobiose · hydrogen bonds · molecular recognition · receptors

- [1] A. J. Thompson, R. P. de Vries, J. C. Paulson, *Curr. Opin. Virol.* **2019**, *34*, 117–129.
- [2] a) J. Balzarini, *Nat. Rev. Microbiol.* **2007**, *5*, 583–597; b) I. Bagdonaitė, H. H. Wandall, *Glycobiology* **2018**, *28*, 443–467; c) E. Keyaerts, L. Vijgen, C. Pannecouque, E. Van Damme, W. Peumans, H. Egberink, J. Balzarini, M. Van Ranst, *Antiviral Res.* **2007**, *75*, 179–187; d) C. A. Mitchell, K. Ramessar, B. R. O’Keefe, *Antiviral Res.* **2017**, *142*, 37–54.
- [3] a) M. F. Bravo, M. A. Lema, M. Marianski, A. B. Braunschweig, *Biochemistry* **2020**, <https://doi.org/10.1021/acs.biochem.0c00732>; b) S. Tommasone, F. Allabush, Y. K. Tagger, J. Norman, M. Köpf, J. H. R. Tucker, P. M. Mendes, *Chem. Soc. Rev.* **2019**, *48*, 5488–5500.
- [4] O. Francesconi, S. Roelens, *ChemBioChem* **2019**, *20*, 1329–1346.
- [5] P. Stanley, H. Schachter, N. Taniguchi in *Essentials of Glycobiology, Chapter 8 N-Glycans*, (Eds: A. Varki, R. D. Cummings, J. D. Esko, et al.), 2nd ed., Cold Spring Harbor, New York, USA, **2009**.
- [6] C. J. Burrell, C. R. Howard, F. A. Murphy in *Fenner and White’s Medical Virology. Chapter 4—Virus Replication*, (Eds: C. J. Burrell, C. R. Howard, F. A. Murphy), 5th ed., Academic Press, London, **2017**, pp. 39–55.
- [7] a) S. C. Gordts, M. Renders, G. Férrir, D. Huskens, E. J. M. Van Damme, W. Peumans, J. Balzarini, D. Schols, *J. Antimicrob. Chemother.* **2015**, *70*, 1674–1685; b) J. Balzarini, K. Van Laethem, S. Hatse, M. Foreyen, W. Peumans, E. Van Damme, D. Schols, *J. Biol. Chem.* **2005**, *280*, 41005–41014.
- [8] A. P. Davis, *Chem. Soc. Rev.* **2020**, *49*, 2531–2545.
- [9] a) E. A. Kataev, C. Muller, *Tetrahedron* **2014**, *70*, 137–167; b) G. V. Oshovsky, D. N. Reinhoudt, W. Verboom, *Angew. Chem. Int. Ed.* **2007**, *46*, 2366–2393; *Angew. Chem.* **2007**, *119*, 2418–2445.
- [10] Selected examples from recent literature: a) D. W. Zhang, A. Martinez, J. P. Dutasta, *Chem. Rev.* **2017**, *117*, 4900–4942; b) T. Hayashi, Y. Ohishi, H. Abe, M. Inouye, *J. Org. Chem.* **2020**, *85*, 1927–1934; c) Y. Ohishi, N. Yamamoto, H. Abe, M. Inouye, *J. Org. Chem.* **2018**, *83*, 5766–5770; d) M. Yamashina, M. Akita, T. Hasegawa, S. Hayashi, M. Yoshizawa, *Sci. Adv.* **2017**, *3*, e1701126; e) X. Schaapkens, E. O. Bobylev, J. H. H. Reek, T. Mooibroek, *Org. Biomol. Chem.* **2020**, *18*, 4734–4738.
- [11] O. Francesconi, M. Martinucci, L. Badii, C. Nativi, S. Roelens, *Chem. Eur. J.* **2018**, *24*, 6828–6836.
- [12] O. Francesconi, F. Cicero, C. Nativi, S. Roelens, *ChemPhysChem* **2020**, *21*, 257–262.
- [13] O. Francesconi, A. Ienco, C. Nativi, S. Roelens, *ChemPlusChem* **2020**, *85*, 1369–1373.
- [14] Examples of adaptive architectures in carbohydrate recognition: a) Y. H. Kim, J. I. Hong, *Angew. Chem. Int. Ed.* **2002**, *41*, 2947–2950; *Angew. Chem.* **2002**, *114*, 3071–3074; b) M. Mazik, *RSC Adv.* **2012**, *2*, 2630–2642; c) P. Mateus, B. Wicher, Y. Ferrand, I. Huc, *Chem. Commun.* **2018**, *54*, 5078–5081; d) Y. Ohishi, K. Masuda, K. Kudo, H. Abe, M. Inouye, *Chem. Eur. J.* **2021**, *27*, 785–793.
- [15] A. Vacca, O. Francesconi, S. Roelens, *Chem. Rec.* **2012**, *12*, 544–566.
- [16] T. J. Mooibroek, J. M. Casas-Solvas, R. L. Harniman, C. M. Renney, T. S. Carter, M. P. Crump, A. P. Davis, *Nat. Chem.* **2016**, *8*, 69–74.
- [17] J. L. Asensio, F. J. Cañada, M. Bruix, A. Rodriguez-Romero, J. Jiménez-Barbero, *Eur. J. Biochem.* **1995**, *230*, 621–633.
- [18] Observed short intermolecular distances to anthracene planes: H-1, 2.64 Å; H-2, 2.86 Å; H-3, 2.69 Å; H-4, 2.81 Å; H'-1, 2.96 Å; H'-2, 2.77 Å; H'-5, 2.67 Å. Me (Ac) to Carbazole plane: 2.92 Å.
- [19] M. C. Fernández-Alonso, D. Díaz, M. Á. Berbis, F. Marcelo, F. J. Cañada, J. Jiménez-Barbero, *Curr. Protein Pept. Sci.* **2012**, *13*, 816–830.

Manuscript received: January 13, 2021

Revised manuscript received: February 26, 2021

Accepted manuscript online: March 5, 2021

Version of record online: April 6, 2021