ARTICLE

OBC
www.rsc.org/obc

Enzymatic synthesis of complex glycosaminotrioses and study of their molecular recognition by hevein domains †

Nuria Aboitiz,[a] F. Javier Cañada,[a] Lucie Hušáková,[b] Marek Kuzma,[b] Vladimír Křen*[b] and Jesús Jiménez-Barbero*[a]

[a] *Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain. E-mail: jjbarbero@icib.csic.es; Fax: +34915360432; Tel: +34918373112*

[b] *Institute of Microbiology, Academy of Sciences of the Czech Republic, VídeČská 1083, 142 20 Prague, Czech Republic. E-mail: kren@biomed.cas.cz*

Received 23rd January 2004, Accepted 30th April 2004
First published as an Advance Article on the web 28th June 2004

Hevein, a protein found in *Hevea brasiliensis,* has a CRD domain, which is known to bind chitin and GlcNAc-containing oligosaccharides. By using NMR and molecular modeling as major tools we have demonstrated that trisaccharides containing GalNAc and ManNAc residues are also recognized by hevein domains. Thus far unknown trisaccharides GlcNAcβ(1⟶4)GlcNAcβ(1⟶4)ManNAc (**1**) and GalNAcβ(1⟶4)GlcNAcβ(1⟶4)ManNAc (**2**) were synthesized with the use of β-*N*-acetylhexosaminidase from *Aspergillus oryzae*. This method is based on the rather unique phenomenon that some fungal β-*N*-acetylhexosaminidases cannot hydrolyze disaccharide GlcNAcβ(1⟶4)ManNAc (**5**) contrary to chitobiose GlcNAcβ(1⟶4)GlcNAc (**4**) that is cleaved and, therefore, cannot be used as an acceptor for further transglycosylation. Both trisaccharides **1** and **2** were prepared by transglycosylation from disaccharidic acceptor **5** in good yields ranging from 35% to 40%. Our observations strongly indicate that the present nature of the modifications of chitotriose (GlcNAcβ(1⟶4)GlcNAcβ(1⟶4)GlcNAc, **3**) at either the non-reducing end (GalNAc instead of GlcNAc) or at the reducing end (ManNAc instead of GlcNAc) do not modify the mode of binding of the trisaccharide to hevein. The association constant values indicate that chitotriose (**3**) binding is better than that of **1** and **2**, and that the binding of **1** (with ManNAc at the reducing end) is favored with respect to that of **2** (with ManNAc at the reducing end with a non-reducing GalNAc moiety).

Introduction

In recent years, it has been shown that the interactions between carbohydrates and proteins mediate a broad range of biological activities, starting from fertilization, embryogenesis, and tissue maturation, and extending to such pathological processes as tumor metastasis.[1] The elucidation of the mechanisms that govern how oligosaccharides are accommodated in the binding sites of lectins, antibodies, and enzymes is currently a topic of major interest.[2] It is obvious that detailed knowledge of the structural features, dynamics, and energetics of the complex when carbohydrates are bound to lectins (non-enzymatic carbohydrate binding proteins) and enzymes is indeed relevant. The concerted use of a variety of biophysical, biochemical and spectroscopic techniques together with the access to synthetically prepared oligosaccharides and analogues, as well as to natural and "designed" protein domains is of paramount importance. X-Ray, NMR, and modeling are among the methods[3] that have been more widely used to access detailed structural and thermodynamic information.

The presence of the oxygen atoms of the hydroxyl groups obviously makes possible their involvement in intermolecular hydrogen bonds to side-chains of polar amino acids within the polypeptide chain. Nevertheless, not only polar forces are involved in carbohydrate recognition. Depending on the stereochemistry of the monomer constituents of the oligosaccharide chain, the presence of a number of rather apolar C–H groups indeed constitutes patches that provide van der Waals, CH–π, and hydrophobic interactions.

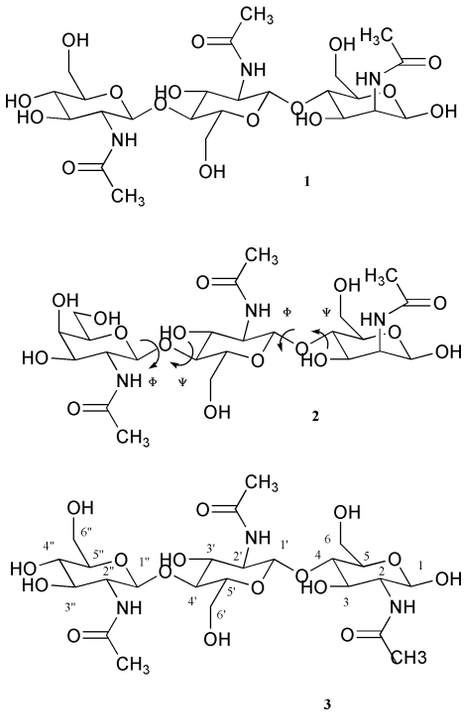
X-Ray crystallography analysis is frequently performed to solve such problems.[4] The use of spectroscopic analysis alone or in combination with computations provides further means to study these interactions.[5] Indeed, the use of fluorescence and NMR spectroscopy with lectins in solution has also been conducive to underscore the already emphasized role of aromatic residues in the binding site of lectins for stacking.[6]

Among the biological processes in which carbohydrates are involved, many plants respond to pathogenic attack by producing defense proteins[7] able to bind reversibly to chitin, a β(1⟶4) linked 2-acetamido-2-deoxy-D-glucopyranose polysaccharide (Scheme 1).

This natural biopolymer is a key structural component of the cell wall of fungi, and of the exoskeleton of invertebrates, such as insects and nematodes. Most of these defense proteins include a common structural motif of 30–43 residues rich in glycines and cysteines in highly conserved positions and organized around a four disulfide core, usually known as a hevein domain or a chitin binding motif.[8] The hevein domain is present in several lectins, like hevein itself and its natural variant pseudohevein, *Urtica dioica* agglutinin (UDA), wheat germ agglutinin (WGA), and Ac-AMP antimicrobial peptides.[9] The same chitin binding motif can also be found in enzymes with antifungal activity, like class I chitinases. This biological activity is probably related to the catalytic properties of the protein. Thus, the fungal growth is probably limited by the degradation of fungal cell walls caused by the hydrolytic action of the enzyme. Surprisingly, small chitin-binding proteins that contain the hevein domain, like hevein itself, or Ac-AMP peptides have been shown to have a remarkable antifungal, and antinutrient activity in insects even though they do not have any known enzymatic activity. Its small size (43

† Electronic supplementary information (ESI) available: Figures: potential energy maps and global minimum conformers of disaccharide entities of **1**–**3**; trajectory Φ/Ψ plots for **1** and **2**; COSY and TOCSY spectra for **1** and **2**; titration points obtained upon addition of increasing amounts of **1** and **2** to hevein; backbone rms deviation from initial NMR structure along solvated MD simulations for complexes of hevein with **1**–**3**. Tables: relative energy and geometric features of possible conformers of disaccharide entities of **1**–**3**; ^{1}H and ^{13}C NMR data. See http://www.rsc.org/suppdata/ob/b4/b401037j/

DOI: 10.1039/b401037j



Scheme 1 A schematic view of **1**, **2** and **3**.

residues) and availability by purification, molecular biology, or peptide synthesis methods makes this domain an excellent model system for the study of carbohydrate recognition by proteins.

The work by Wright[10] has indeed become the basis for further studies on proteins displaying this domain, including hevein itself,[11] both in the solid state and in solution. The hevein 3D-structure has independently been solved in the solid state by X-ray studies at 2.8 Å resolution,[12] and in solution by NMR methods both in water[13] and in dioxane/water.[14]

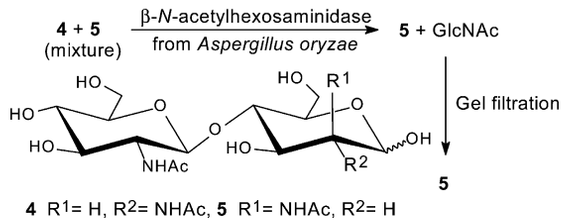
According to the X-ray and NMR studies, the aromatic residues at relative positions 21, 23 and 30 in these domains play an important role in carbohydrate binding, stabilizing the complexes by means of stacking interactions and van der Waals contacts. Additionally, Ser-19 is involved in a hydrogen bonding interaction with the carbonyl group of the acetamide moiety of the GlcNAc residue stacking on the aromatic ring of residue 23. As a final key interaction, the hydroxyl group of Tyr 30 also provides additional hydrogen bonding with OH-3 of the same sugar moiety. According to this topology, it has been largely assumed that hevein domains interact specifically with GlcNAc-containing oligosaccharides.[15] Nevertheless, by using NMR and molecular modeling as major tools, we will show in this paper that trisaccharides containing GalNAc and ManNAc residues are also recognized by hevein domains. A schematic view of their structures is also given in Scheme 1.

To our knowledge, oligosaccharides combining in their molecule all three common *N*-acetylhexosamines linked by β(1⟶4) linkages (analogously to chitine) had never been prepared in the past.

Results and discussion

Enzymatic synthesis of trisaccharides 1 and 2

Trisaccharide **1** may be prepared from chitotriose (**3**) using Lobry de Bruyn–van Ekenstein epimerization in saturated $Ca(OH)_2$ solution.[16] We have used analogous methodology for the preparation of GlcNAcβ(1⟶4)ManNAc (**5**) from *N,N′*-diacetylchitobiose (**4**).[16] However, the major problem is the separation of the product **5** with *manno* configuration (20% yield) from unreacted starting material **4**. Higher yields of the *manno*-derivatives cannot be achieved due to the limit of the thermodynamic equilibrium. The separation problem was recently solved in our laboratories by using selective enzymatic discrimination using β-*N*-acetylhexosaminidase from *Aspergillus oryzae*. The enzyme discriminates the mixture of disaccharides **4** and **5**: *N,N′*-diacetylchitobiose (**4**) was selectively hydrolyzed by β-*N*-acetylhexosaminidase, whereas its C-2 epimer (**5**) was completely resistant to the enzyme hydrolysis. Resulting disaccharide **5** can be easily separated from the GlcNAc formed by gel filtration (Scheme 2).[17]



Scheme 2 Selective removal of GlcNAcβ(1⟶4)GlcNAc (**4**) from the epimerization mixture by β-*N*-acetylhexosaminidase from *Aspergillus oryzae*.

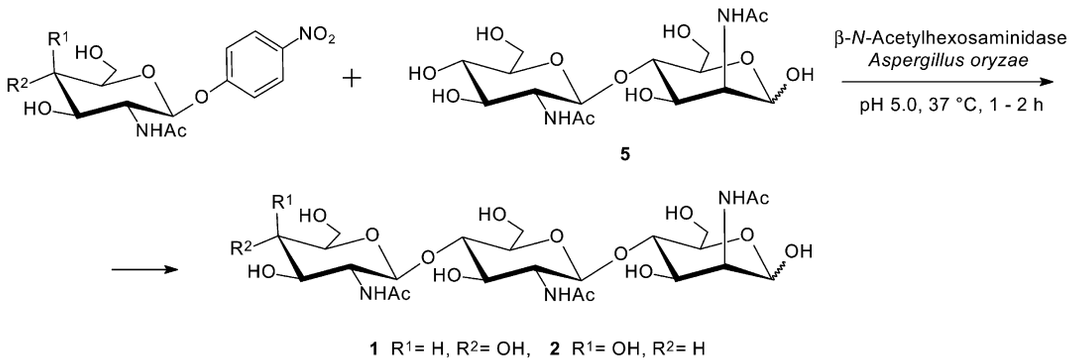
This method cannot be, however, used for analogous separation of the mixture of **1** and **3** because the enzyme hydrolyses in both trisaccharides nonreducing GlcNAc forming a mixture of **4** and **5**, which is eventually transformed into the above mentioned mixture of GlcNAc and **5**. Fungal β-*N*-acetylhexosaminidases, especially that from *A. oryzae* is, nevertheless, able to transfer β-GlcNAc and β-GalNAc residues onto various acceptors in rather good yields.[18,19] In the case of Glc- or GlcNAc-terminal acceptors the enzyme prefers the C-4 OH group for the attachment of an additional sugar.[19] *p*-Nitrophenyl 2-acetamido-2-deoxy-β-D-glucopyranoside and *p*-nitrophenyl 2-acetamido-2-deoxy-β-D-galactopyranoside are common glycosyl donors in such reactions (Scheme 3).

If the *N,N′*-diacetylchitobiose (**4**) were used for further extension with β-*N*-acetylhexosaminidase-catalyzed glycosylation using *p*NP-βGlcNAc or *p*NP-βGalNAc, the enzyme would hydrolyze in parallel both donor and acceptor giving rise to a rather complex reaction mixture containing only little trisaccharidic (moreover, not uniform) fraction. However, when employing an advantage of **5**, which is resistant to enzymatic cleavage but could serve as an acceptor, we can prepare both trisaccharides **1** (using *p*NP-βGlcNAc) and **2** (*p*NP-βGalNAc) in good yields (35–40%). The reaction mixture is clean, containing only single trisaccharide (product) and disaccharide (**5**) and some monosaccharidic byproduct. The product can be isolated simply using gel filtration and, in parallel, also unreacted starting material can be regenerated. Of course, this single step reaction affords directly the required products in their free forms, and no tedious protection/deprotection steps or extensive chromatographies with organic solvents are required.

The conformations of trisaccharides 1 and 2 in the free state

A conformational study of both trisaccharides **1** and **2** was carried out by using molecular mechanics and dynamics calculations, and compared to that of **3**.

In a first approximation, the potential energy Φ/Ψ surfaces for the three disaccharides, of which **1** and **2** are composed [GlcNAcβ(1⟶4)GlcNAc (A1), GalNAcβ(1⟶4)GlcNAc (A2), GlcNAcβ(1⟶4)ManNAc (B)], were calculated with the MM3* force field[20] as integrated in MACROMODEL[21] (see electronic supplementary information†). Nomenclature of Φ and Ψ is described in the experimental section. This force field was chosen, since it has provided a fair agreement between experimental and expected NMR data for chitooligosaccharides.[22] A well-established methodology[23–25] was applied, as described in the experimental section, that permitted verification that the



Scheme 3 Preparation of the trisaccharides **1** and **2** using β-*N*-acetylhexosaminidase from *Aspergillus oryzae*.

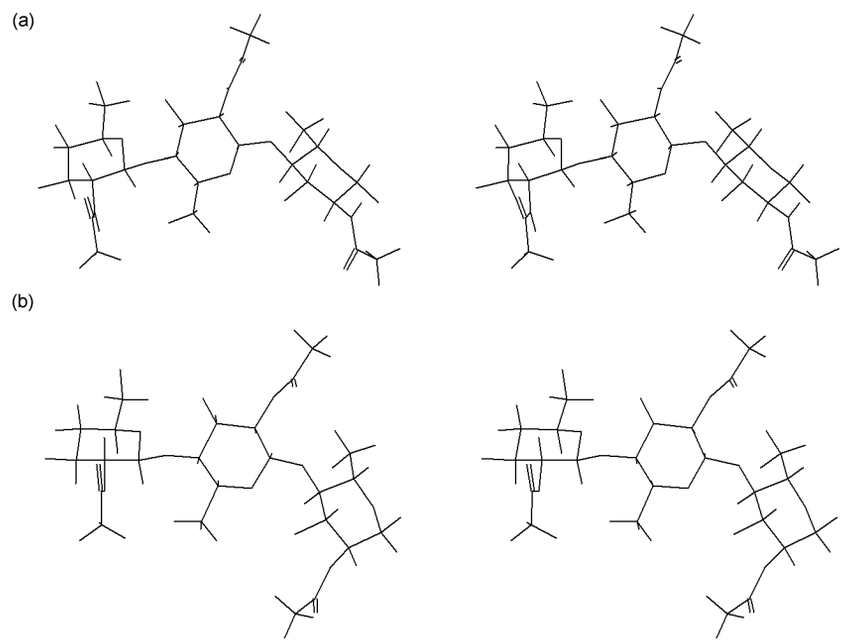


Fig. 1 Stereoscopic views of the global minimum conformer (double *synΦ*/*synΨ*) of trisaccharides **1** (top) and **2** (bottom) according to MM3* (GB/SA) calculations.

potential energy surfaces for the three disaccharide entities are very similar, as shown in the electronic supplementary information. † In all cases, a very major and two minor conformational families are predicted (Table S1 in the electronic supplementary information †).

According to the MM3* force field, each of these families shows a certain degree of motion around the corresponding energy minimum. About 98% of the population of A1, A2, and B is located in the minimum *synΦ*/*synΨ* (a), which corresponds to dihedral angles of $\Phi = 50° \pm 20°$ and $\Psi = 0° \pm 20°$. Minimum *synΦ*/*antiΨ* ($\Phi = 45° \pm 10°$ and $\Psi = 175° \pm 10°$) shows *ca.* 1% of the population. An additional local minimum *antiΦ*/*synΨ* ($\Phi = 180° \pm 20°$ and $\Psi = 0° \pm 20°$) is calculated to account for *ca.* 1% of the population for A1, A2, and B, independently of the nature of the sugars flanking the β(1⟶4) glycosidic linkage. Thus, as expected, *exo*-anomeric conformers around both glycosidic angles are always present.

The conformational entropy[26] was calculated from the probability distribution of conformers of **1** and **2**, using the MM3* force field at 300 K, and amounted to *ca.* 8.8 kJ mol^{-1} per glycosidic linkage independently of the chemical nature (GlcNAc, ManNAc, GalNAc) of the constituent sugars.[23,25,27,28] Taking into account the two linkages for both **1** and **2**, although only qualitative, this number indicates that the freezing of these ligands upon binding could generate a maximum entropy loss of about 17 kJ mol^{-1}.

As second step, the conformational behavior of the two trisaccharides **1**,**2** was analyzed by molecular dynamics. After the generation of nine trisaccharide structures from the local minima of the constituent disaccharide entities, molecular mechanics calculations with the MM3* program and the GB/SA solvent model for water were performed. The MM3* values results clearly indicate that the double *synΦ*/*synΨ* conformers at the glycosidic linkages are much more stable than the other conformers (Fig. 1). Indeed, according to a Boltzmann distribution from the relative energy values, they amount to more than 96% of the population. In order to test the conformational stability of these minima, independent 5 ns MD-simulations (see electronic supplementary information †) were performed on both trisaccharides, starting with the double *synΦ*/*synΨ* conformers at either glycosidic linkage.

The resulting trajectories showed basically no interconversions to the *anti* regions during the 5 ns simulation time. A visual inspection permits us to state that they resemble the potential energy surfaces obtained for the disaccharide entities. However, there is a fair amount of conformational averaging around the *synΦ*/*synΨ* conformers. Less than 1% of the *anti* rotamers are found. The MM3*-based MD predict a higher flexibility around Ψ than around Φ angles, as expected for the existence of the *exo*-anomeric effect.

The chemical shifts in D_2O (and H_2O to detect the amide protons) of **1** and **2** are listed in the experimental section. The assignment of the resonances was made through a combination of COSY, TOCSY, NOESY/ROESY, and HSQC experiments (see electronic supplementary information †). Intense NOEs were obtained for H1″–H4′ and H1′–H4 proton pairs and observable NOEs were detected for H1″–H6′ab and H1′–H6ab proton pairs, as expected for double *synΦ*/*synΨ*

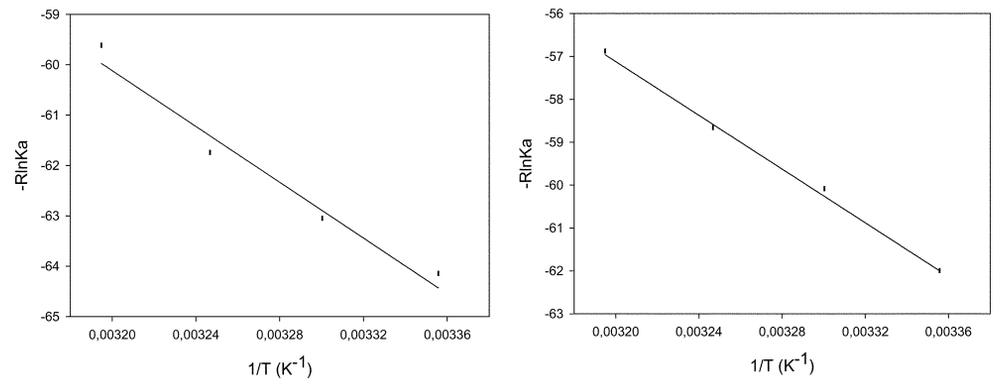


Fig. 2 Van't Hoff plots for the binding of the trisaccharides (left: **1**, right: **2**) to hevein. Although the van't Hoff data may be approximate, the slopes indicate that the process is enthalpy driven and that entropy opposes binding, as also deduced for regular chitooligosaccharide binding to hevein domains.

Table 1 Association constant data, K_a (M^{-1}), for trisaccharides, as deduced from titration NMR experiments. The variations of different ^1H resonances of hevein were followed upon addition of increasing amounts of **1** and **2** to the NMR tube containing a constant concentration of hevein.

T/K	**1**	**2**
298	2253 ± 232	1742 ± 243
303	1975 ± 213	1384 ± 203
308	1688 ± 163	1166 ± 176
313	1307 ± 113	941 ± 108

conformers (Table S1 in the electronic supplementary information †).

Although their quantitative analysis was obscured by severe overlapping, giving the analogy to **3** and to other reported β(1⟶4) linked saccharides,[22,25,27] it may be safely assumed that the NMR data are in agreement with the MM3* predictions and that the double *synΦ*/*synΨ* geometries are the major ones existing in solution. Thus, these geometries were considered as input for the derivation of the 3D structures of their complexes to hevein.

The binding to hevein

A first estimation of the binding constant could be obtained by simple 1D NMR measurements. Thus, the binding of **1** and **2** to hevein was monitored by recording ^1H-NMR spectra of a series of samples with increasing sugar concentration, in which the concentration of protein during the experiments was held constant. In both cases, the observed effects on the chemical shifts and line broadening indicate that the interaction is basically fast on the chemical shift NMR time scale. The signals for W21, W23 and Ser19 are particularly affected by the addition of both sugars **1** and **2**, in close resemblance to the data of binding of chitin fragment **3** to hevein.[13,15] These observations indicate that the present nature of the modifications at either the non-reducing end (GalNAc instead of GlcNAc) or at the reducing end (ManNAc instead of GlcNAc) induces no substantial modification of the mode of binding of the trisaccharide to hevein.

The association constant values for **1** and **2** in comparison to those of the basic $(GlcNAc)_3$ moiety (**3**) are given in Table 1, while some examples of the type of spectra acquired are given in the electronic supplementary information. †

The association constant values shown in Table 1 indicate that chitotriose binding[13,15] is more efficient than that of **1** and **2**, and that the binding of **1** (with ManNAc at the reducing end) is favored with respect to that of **2** (with MaNAc at the reducing end and a non-reducing GalNAc moiety).

In a further step, a van't Hoff of the NMR-based association constants as a function of temperature was performed to give the approximated equilibrium thermodynamic parameters, ΔH^0 and ΔS^0 (Fig. 2). It is noteworthy to consider that although the van't Hoff data may be only approximated, previous studies from our group for a variety of chitin oligosaccharide/hevein domain systems[13,15] have permitted us to demonstrate that the obtained NMR values only differ marginally (10% at most) from those obtained by titration microcalorimetry.

In all the investigated cases the entropy of binding ΔS^0 was found to be negative, as observed for a variety of GlcNAc-containing oligosaccharides interacting with hevein, pseudo-hevein, WGA, and UDA as well. For instance, for chitotriose binding to hevein the corresponding values[13,15] are ΔH^0 of -36.4 kJ mol^{-1} and ΔS^0 -45.1 J mol^{-1} K^{-1}. In the present cases, the enthalpy values are smaller for **1** and **2** and amount to -27.7 and -31.3 kJ mol^{-1}, respectively, while their entropy losses are smaller than or equal to that of **3**, reaching -28.5 and -45.1 J mol^{-1} K^{-1}, respectively. Thus, in spite of a relatively large enthalpy of binding, the dissociation constants measured are in the millimolar range, since the entropy of binding highly opposes the association. Although the origin of this enthalpy–entropy compensation phenomenon remains an open question,[27,29] it has been reported[30] for this magnitude and sign of ΔS and ΔH that hydrogen bonding and van der Waals forces should be the most important factors that stabilize the complex. The negative entropy of binding observed could arise from rigidification of the sugar and/or the protein lateral chains[29] or from reorganization of the water structure.[29]

As mentioned above, the maximum loss of conformational entropy by freezing of these ligands upon binding to hevein could generate a maximum entropy loss of about 17 kJ mol^{-1}, much higher than that experimentally observed (between 8.5 and 13.5 kJ mol^{-1}). Although the observed entropy values are only approximate, the difference may indicate the existence of saccharide motion even when bound to the protein.

The models of the complexes of 1 and 2 with hevein

From the structural viewpoint, the 3D structures of the complexes of hevein to **1** and **2** were obtained by using molecular modeling methods, supported by our previous NMR studies, which demonstrate that two different three-dimensional structures of the complexes of hevein domains to $(GlcNAc)_3$ are possible. The only difference between the two possible structures for each complex resides in the relative position of the trisaccharide with respect to the binding site: the non-reducing end occupies different protein subsites. In structure A (Fig. 3A), the acetamido methyl group of the non-reducing end shows non-polar contacts with two aromatic residues: Tyr30 and Trp23, and, in addition, there are important hydrogen bonds, which confer stability to the complex: one between the terminal non-reducing sugar acetamide group and Ser19 and a second one involving C3–OH and Tyr30. Two key CH–π interactions are observed: one between the β-face of the terminal non-reducing sugar moiety and the plane of the aromatic ring

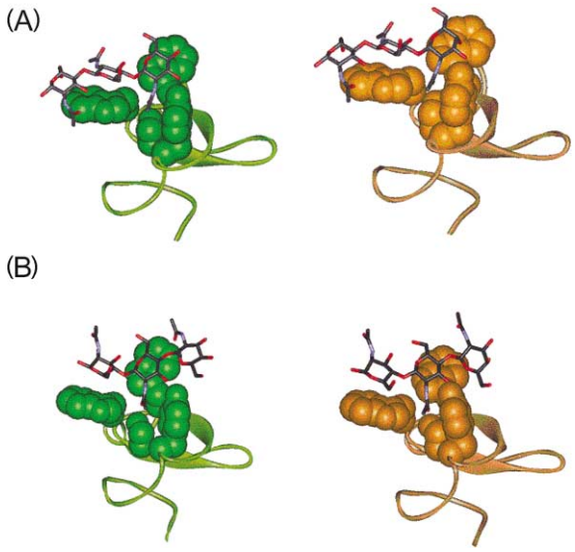


Fig. 3 A: Complexes of type A are NOT viable. In these views it can be observed that the NAc chain of the reducing ManNAc moiety of both trisaccharides (left: **1**, right: **2**) make steric clashes with the aromatic ring Trp 23. B: Complexes of type B are viable. In these views it can be observed that both Trp aromatic rings make stabilizing contacts with the central GlcNAc and the reducing ManNAc moieties of both trisaccharides (left: **1**, right: **2**).

of Trp23 and a second one between the less polar face of the central GlcNAc moiety and the plane of the aromatic ring of Trp21. In a second complex (complex B, Fig. 3B), the central sugar unit interacts with Ser19, Trp23 and Tyr30, while the reducing residue makes contacts with Trp21. Additional van der Waals contacts between the non-reducing end and Cys24 take place.

Therefore, compounds **1** and **2** were docked on the hevein binding site using the topologies of both types of complexes A and B as starting structures. The docking and minimization results showed that only one complex (B-type) is indeed possible (Fig. 3B). GlcNAc is required as the central unit to provide key interactions with Trp 23. The ManNAc residue provides the additional interactions with Trp21, and the non-reducing GalNAc moiety provides minor additional favorable contacts with the polypeptide, especially with Cys24. In contrast, complexes of type A are not viable (Fig. 3A). In these complexes, for which the non-reducing fragment would provide the interactions with Trp 23, and the central GlcNAc moiety those with Trp21, the reducing ManNAc ring would make bad contacts with the extended face of Trp21, due to the different orientation of the acetamide lateral chain, pointing towards the aromatic ring.[13,15] In addition, the axial orientation of the OH-4 of GalNAc would block the key CH–π interaction with the aromatic surface of Trp23. The resulting geometries were submitted to an energy minimization process, followed by solvated MD simulations.

In particular, the solvated molecular dynamics (4.5 ns) studies were carried out with the AMBER 5.0 program,[31] using the protein coordinates from the published NMR structures of hevein/$(GlcNAc)_3$ and the MM3* coordinates of the global minima of **1** and **2** as starting geometries to get the complexes depicted in Fig. 3B.

During the MD simulation, the conformational space accessible to the protein is basically identical in all cases (Fig. 4). The backbone is fairly well defined and it maintains the same topology, with backbone rmsd to the starting structure during the whole simulation smaller than 1 Å (see electronic supplementary information †). In addition, the orientation of residues Ser19, Trp21, Trp23, and Tyr30 in the recognition site of hevein obtained after the modeling protocol is also very similar in the free state and in the complex with chitooligosaccharides, according to our previous NMR studies.[13,15] All

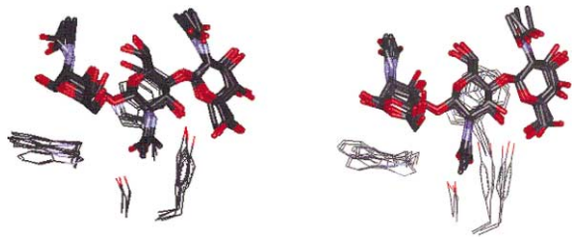


Fig. 4 Superimposition of 5 structures obtained from a molecular dynamics simulation (4.5 ns) in the presence of explicit water molecules. Models of the trisaccharide (left: **1** right: **2**) sitting on the binding site of hevein. The topology of the binding site and of the complexed sugar is very similar during the whole simulation.

these observations strongly suggest that very minor changes are indeed required to accommodate the sugar moiety in the binding site of hevein domains. Regarding the role of water molecules, no special indication of long resident times of specific water molecules was deduced from the radial distribution computed for the key Ser19 and Tyr30 hydroxyl groups. These hydroxyls keep the hydrogen bonding pattern to the *N*-Ac carbonyl and the 3-OH of the central GlcNAc unit, respectively, during the complete simulation.

From the perspective of the sugars, the glycosidic linkages of the trisaccharides present the regular Φ/Ψ angles for β(1⟶4) linked oligosaccharides (double *syn*-$\Phi\Psi$) throughout the MD simulation. The double sugar–aromatic stacking with the central and reducing ends on top of the Trp23 and Trp21 aromatic rings, make that double *syn*-$\Phi\Psi$ conformation particularly stable. This carbohydrate:aromatic CH–π interaction is sensitive to the configuration of the anomeric center,[13,15] and is maximized when the disaccharide adopts a double *syn* geometry. Nevertheless, still some motion (Fig. 4) remains for the sugar moieties in the binding site (rmsd *ca.* 0.6 Å)

In conclusion, the solution structures of hevein complexed to **1** and **2** (modeled in this work) are very similar to those obtained for the B-complexes of hevein with **3** by NMR (backbone rmsd smaller than 1 Å).

Regarding the experimental NMR data, the observed binding affinity of **1** and **2** towards hevein is *ca.* one order of magnitude smaller (more than 4.18 kJ mol^{-1} in binding energy) than that of $(GlcNAc)_3$. In fact, their association constants are similar to that found for disaccharide binding in the hevein/$(GlcNAc)_2$ complex.[13] The observed increase of affinity for hevein when passing from the $(GlcNAc)_2$ to the $(GlcNAc)_3$ complex has been explained by the combined effect of the two binding modes existing for trisaccharide binding and by the additional stacking interactions provided by the reducing GlcNAc residue in A-type complexes with the extended face of Trp23. As detailed above, the change in stereochemistry of the *N*-Ac moiety of the terminal ManNAc with the concomitants steric clashes with Trp23 (Fig. 3A) unit makes impossible the existence of these A-type complexes for both **1** and **2**. Therefore, in these analogues there are not any additional stabilizing interactions in the complexes. Nevertheless, these computational data and the NMR data reported herein indicate that hevein domains may bind saccharide moieties different from GlcNAc provided that the basic van der Waals and hydrogen bonding interactions are kept in the complexes.

The material and methodology presented here can serve, not only for the probing of the hevein CRD, but also for the detailed study of the CRD of other lectins with affinity towards aminosugards, and for the design of optimal ligand usable for glycodrug synthesis.

Experimental

NMR spectroscopy

The NMR experiments for assignment of the sugar protons were performed at 500 MHz on Bruker AVANCE spectro-

meters, between 299 and 318 K, and sugar concentrations between 1 and 2 mM.

The corresponding compound was dissolved in D_2O and the solution was degassed by passing argon. COSY, TOCSY (80 ms, mixing time), and ge-HSQC experiments were performed using standard sequences at temperatures between 298 and 310 K. 2D T-ROESY experiments were performed with mixing times of 300, 400 and 500 ms. The strength of the 180 pulses during the spin lock period was attenuated four times with respect to that of the 90 hard pulses (between 7.2 and 7.5 μs).

NMR spectra of **2** and **1a** were measured on a Varian Inova-400 spectrometer (399.89 MHz and 100.55 MHz, respectively) in D_2O or $CDCl_3$ at 303 K. Residual signal of $CDCl_3$ was used as an internal standard (δ_H 7.265, δ_C 77.00); D_2O spectra were referenced to acetone (δ_H 2.030, δ_C 30.50). ^{1}H NMR, COSY, TOCSY, HMQC, HMBC, and ROESY spectra were measured using standard manufacturers' software (Varian Inc., USA). Selective 1D-TOCSY was measured with sequence published by Uhrín *et al.*[32] ^{1}H NMR spectra were zero filled to fourfold the number of data points and multiplied by the window function (two-parameter double-exponential Lorentz–Gauss function) before Fourier transformation to improve resolution. Protons were assigned by COSY and TOCSY and the assignment was transferred to carbons by HMQC (Table S2 in electronic supplementary information†). Digital resolution allowed us to report chemical shifts of protons to three and coupling constants to one decimal place. Carbon chemical shifts were read out from HMQC (protonated carbons) and are reported to one decimal place. The peracetylated form of compound **1** was used for structure elucidation to obtain better signal dispersion.

Mass spectrometric analysis

Mass spectra were measured on a matrix-assisted laser desorption/ionization reflectron time-of-flight MALDI-TOF mass spectrometer BIFLEX (Bruker-Franzen, Bremen, Germany) equipped with a nitrogen laser (337 nm) and griddles delayed extraction ion source. Ion acceleration voltage was 19 kV and the reflectron voltage was set to 20 kV. Spectra were calibrated externally using the monoisotopic $[M + H]^+$ ion of peptide standards human angiotensin I (Sigma). A saturated solution of α-cyano-4-hydroxy-cinnamic acid in 50% MeCN/0.3% acetic acid was used as a MALDI matrix. 1 μl of matrix solution was mixed with 1 μl of the sample on the target and the droplet was allowed to dry at ambient temperature.

Molecular mechanics and dynamics calculations

The molecular mechanics calculations were performed on both trisaccharides, and their disaccharide components using the MM3* force field as implemented in MACROMODEL 5.5 and the GB/SA solvent model for water. Potential energy maps were calculated as described.[25,26]

The non-reducing ends are double primed and the central units are primed (Scheme 1). Thus, the glycosidic torsion angles are defined as Φ'': GlcNAc-H1″–GlcNAc-C1″–GlcNAc-O1″–GlcNAc-C4′, Ψ'' GlcNAc-C1″–GlcNAc-O1″–GlcNAc-C4′–GlcNAc-H4′, Φ' GlcNAc-H1′–GlcNAc-C1′–GlcNAc-O1′–ManNAc-C4, Ψ' GlcNAc-C1′–GlcNAc-O1′–ManNAc-C4–ManNAc-H4 for **1** and GalNAc-H1″–GalNAc-C1″–GalNAc-O1″–GlcNAc-C4′, Ψ'' GalNAc-C1″–GalNAc-O1″–GlcNAc-C4′–GlcNAc-H4′, Φ' GlcNAc-H1′–GlcNAc-C1′– GlcNAc-O1′–ManNAc-C4, Ψ' GlcNAc-C1′–GlcNAc-O1′–ManNAc-C4–ManNAc-H4, for **2**.

MD calculations were also performed with this force field for **1** and **2**. In particular, two independent 5 ns unrestrained MD simulations were run starting from the combination of the global minima of the constituent disaccharide entities.

The MD calculations for the complexes were done using the AMBER 5.0[31] force field. Starting glycosidic torsion angles were taken from the MM3* calculations of **1** and **2** and were slightly modified to fit those described for the NMR-based hevein/oligosaccharide complexes. The atomic charges for the disaccharides in the MD simulations of the complexes were AMBER charges. The complexes were immersed into a box of 3017 water molecules in order to obtain perfect solvation. Cutoff for nonbonding interactions was set to 11.0 Å.

For the molecular dynamics simulations (MD), the structures were constructed using the X-LEAP program.[33] All the MD simulations were carried out using the Sander module within the AMBER 5.0 package. Therefore, the four starting complex geometries were submitted to energy minimization using 600 conjugate gradient iterations after 10 cycles of steepest descent. The two structures corresponding to A-type complexes were not viable. Subsequent molecular dynamics were performed at a constant pressure and temperature, using the Berendsen coupling algorithm for the later, with a time step of 2 fs. Structures were recorded every 0.5 ps for a total calculation time of 4520 and 4840 ps for complexes of **1** and **2** respectively.

NMR titration experiments

The binding of both carbohydrates to hevein was monitored by recording mono-dimensional 500 MHz ^{1}H-NMR spectra of a series of samples with variable sugar concentration (ten different concentrations). The concentration of the protein during the experiments was kept constant (0.3 mM). The hevein sample was prepared by dissolving the lyophilized protein in a volume of 1.0 mL of buffer (85 : 15, ^1H_2O : ^2H_2O, 100 mM sodium chloride, 10 mM sodium phosphate, pH 5.6). The concentration of hevein was calculated from its UV absorbance at 280 nm. The 1D ^{1}H-NMR spectrum for the sample with the highest ligand/protein ratio was recorded by dissolving the corresponding sugar (15 mM) in 0.5 mL of the hevein-containing solution described above. The sample with the other 0.5 mL of this hevein-containing solution was used to obtain the ^{1}H-NMR chemical shifts of the sugar-free protein sample (δ free). The titration curve was established by adding small aliquots of the highest ligand/protein ratio sample to the ligand-free protein sample as previously described.[13] Thermodynamic equilibrium parameters ΔS and ΔH for the hevein–trisaccharide interaction were determined from van't Hoff plots,[13,15] in which the affinity constants were assessed at 25, 30, 35, 40, and 45 °C. Since the use of van't Hoff plots implies the approximation that the heat capacity does not depend on temperature, these results should be considered only as semiquantitative.

β-*N*-Acetylhexosaminidases

Extracellular β-*N*-acetylhexosaminidases (EC 3.2.1.52) from *Aspergillus oryzae* CCF 1066 and from other fungal strains tested originated from the Library of fungal glycosidases of the Laboratory of Biotransformation (Prague) and were prepared by the cultivation of the respective fungi as described previously.[34–37] The producing strain is deposited with the Czech Collection of Fungi (CCF) at the Department of Botany of the Charles University, Prague.

The library of β-*N*-acetylhexosaminidases was screened for the synthesis of **1** and **2**. The enzymes from *A. sojae* CCF 3060, *A. tamarii* CCF 1665, *Penicillium brasilianum* CCF 2155, *P. oxalicum* CCF 2430, *P. funiculosum* CCF 2984, *P. multicolor* CCF 2244, *Talaromyces flavus* CCF 2686 and jack beans (Sigma-Aldrich) hydrolyzed both types of disaccharides **4** and **5** and, therefore, they were not suitable for the synthesis. The enzymes from *A. terreus* CCF 2539, *A flavus* CCF 3056, bovine kidney (Sigma-Aldrich) and bovine epididymis (Sigma-Aldrich) did not hydrolyze the acceptor **5**, but the yields of **1** were rather low and often side products were formed (data not shown). The best yields and clean reactions were achieved with

the enzyme from *A. oryzae* CCF 1066, which was eventually used for all syntheses. Moreover, this enzyme was recently cloned by us[38] and it may be produced in large quantities.

Synthesis of GlcNAcβ(1⟶4)GlcNAcβ(1⟶4)ManNAc (1)

GlcNAcβ(1⟶4)ManNAc (**5**) prepared as described in refs. 16,17 (48.4 mg, 0.12 mmol) and *p*NP-βGlcNAc (24.3 mg, 0.071 mmol) were dissolved in a mixture of citrate–phosphate buffer (50 mM, pH 5.0, 1 mL) and acetonitrile (0.1 mL). β-*N*-Acetylhexosaminidase from *Aspergillus oryzae* CCF 1066 (7 U) was added and the mixture was incubated at 37 °C for 50 min. The reaction was monitored by TLC using silica gel 60 GF_{254} plates (Merck) with the solvent system 2-propanol/water/28% ammonia (7/2/1, v/v). The spots were visualized by charring with 5% H_2SO_4 in ethanol. The reaction was stopped by heating (100 °C for 10 min) and, after removing the *p*-nitrophenol liberated (2 × 1 mL of Et_2O), the mixture was fractionated by gel filtration (BioGel P2, 2.6 × 80 cm, flow rate 11 mL h^{-1}, eluted with H_2O). The isolated yield of the title compound (**1**) was 16.1 mg (36% related to *p*NP-βGlcNAc), $[a]_D^{25} = -19.1$ (*c* 0.465 in water). The compound was peracetylated to obtain clearer NMR spectra by the standard procedure (Ac_2O/py, r.t., 1 day) and the peracetate **1a** was purified by flash chromatography on silica gel with EtOAc/MeOH (95 : 7), yield of **1a** was 12.5 mg (50.6%). The anomeric configuration of mannose units in **1a** was determined from direct coupling constants[39] $J_{C\text{-}1,\ H\text{-}1}$ (177 and 170 Hz, respectively) observed in coupled HMQC. The (1⟶4) linkages in **1a** were inferred from chemical shifts of H-4 and H-4′ showing that these positions are not acetylated and downfield resonating C-4 and C-4′ (with respect to parent ManNAc and GlcNAc), experiencing a glycosylation shift. MALDI-TOF MS: *m*/*z* 986.4 $[M + Na]^+$ (anal. calcd. for $C_{40}H_{57}N_3NaO_{24}$ 986.33); NMR ^1H and ^{13}C—see Table S2 in electronic supplementary information. †

Synthesis of GalNAcβ(1⟶4)GlcNAcβ(1⟶4)ManNAc (2)

Disaccharide (**5**) (45.0 mg, 0.11 mmol) and *p*NP-βGalNAc (25.1 mg, 0.074 mmol) were dissolved in a mixture of citrate–phosphate buffer (50 mM, pH 5.0, 1 mL) and acetonitrile (0.1 mL). β-*N*-Acetylhexosaminidase from *A. oryzae* CCF 1066 (10 U) was added and the mixture was incubated at 37 °C for 2 h. The reaction was stopped by heating (100 °C for 10 min) and, after removal of the *p*-nitrophenol (2 × 1 mL of Et_2O), the mixture was fractionated by gel filtration (BioGel P2, 2.6 × 80 cm, flow rate 12 mL h^{-1}, eluted with H_2O). The isolated yield of the title compound (**2**) was 19.1 mg (41% relative to *p*NP-βGalNAc). The linkages of sugar units in compound **2** were confirmed by inter-residue heteronuclear couplings observed in HMBC. MALDI-TOF MS: *m*/*z* 650.2 $[M + Na]^+$ (Anal. calcd. for $C_{24}H_{41}N_3NaO_{16}$ 650.23); $[a]_D^{25} = +23.2$ (*c* 0.006 in water). NMR ^1H and ^{13}C—see Table S2 in electronic supplementary information. †

Abbreviations

GlcNAc: 2-acetamido-2-deoxy-D-glucopyranose; ManNAc: 2-acetamido-2-deoxy-D-mannopyranose; *p*NP-βGlcNAc: *p*-nitrophenyl 2-acetamido-2-deoxy-β-D-glucopyranoside; *p*NP-βGalNAc: *p*-nitrophenyl 2-acetamido-2-deoxy-β-D-galactopyranoside; GlcNAcβ(1⟶4)GlcNAc: *O*-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-(1⟶4)-2-acetamido-2-deoxy-D-glucopyranose; GlcNAcβ(1⟶4)ManNAc: *O*-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-(1⟶4)-2-acetamido-2-deoxy-D-mannopyranose; GalNAcβ(1⟶4)GlcNAc: *O*-β-(2-acetamido-2-deoxy-β-D-galactopyranosyl)-(1⟶4)-2-acetamido-2-deoxy-D-glucopyranose; $(GlcNAc)_2$: *O*-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-(1⟶4)-2-acetamido-2-deoxy-D-glucopyranose; $(GlcNAc)_3$: *O*-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-(1⟶4)-*O*-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-(1⟶4)-2-acetamido-2-deoxy-D-glucopyranose; GalNAcβ(1⟶4)GlcNAcβ(1⟶4)ManNAc: *O*-(2-acetamido-2-deoxy-β-D-galactopyranosyl)-(1⟶4)-*O*-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-(1⟶4)-2-acetamido-2-deoxy-D-mannopyranose; GlcNAcβ(1⟶4)GlcNAcβ(1⟶4)ManNAc: *O*-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-(1⟶4)-*O*-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-(1⟶4)-2-acetamido-2-deoxy-D-mannopyranose.

Acknowledgements

Support by bilateral project CSIC–AV ČR (J.J.B.–V.K.), by COST D25 (MŠMT OC 25.002) and GAČR (203/04/1045) is gratefully acknowledged. Financial support by DGICYT (Ministry of Science and Technology of Spain) is also acknowledged (Grants BQU2000-1501-C01 and BQU2003-0335-C01). Nuria Aboitiz thanks Comunidad de Madrid for funding. Dr. P. Halada from the Inst. Microbiol, Prague is thanked for the MS spectra measurement.

References

1 *Glycosciences, Status and Perspectives*, ed. H. J. Gabius and S. Gabius, Chapman & Hall, London–Weinheim, 1997.
2 M. R. Wormald, A. J. Petrescu, Y. L. Pao, A. Glithero, T. Elliott and R. A. Dwek, *Chem. Rev.*, 2002, **102**, 371–386.
3 For a current view of the state of the art of carbohydrate NMR, see J. Jiménez-Barbero and T. Peters, *NMR Spectroscopy of Glycoconjugates.*, Wiley-VCH, Weinheim, 2002.
4 (*a*) J. H. Naismith and R. A. Field, *J. Biol. Chem.*, 1996, **271**, 972–976; (*b*) J. Bouckaert, T. W. Hamelryck and L. Wyns and R. Loris, *J. Biol. Chem.*, 1999, **274**, 29188–29195; (*c*) W. I. Weis and K. Drickamer, *Annu. Rev. Biochem.*, 1996, **65**, 441–473.
5 H. C. Siebert, R. Adar, R. Arango, M. Burchert, H. Kaltner, G. Kayser, E. Tajkhorshid, C. W. Lieth, R. Kaptein, N. Sharon, J. F. G. Vliegenthart and H. J. Gabius, *Eur. J. Biochem.*, 1997, **249**, 27–38.
6 H. C. Siebert, C. W. von der Lieth, R. Kaptein, J. J. Beintema, K. Dijkstra, N. van Nuland, U. M. S. Soedjanaatmadja, A. Rice, J. F. G. Vliegenthart, C. S. Wright and H. J. Gabius, *Proteins: Struct., Funct., Genet.*, 1997, **28**, 268–284.
7 W. J. Peumanns and E. J. M. van Damme, *Histochem. J.*, 1995, **27**(4), 253–271.
8 (*a*) N. V. Raikhel, H. I. Lee and W. F. Broekaert, *Annu. Rev. Plant. Physiol.*, 1993, **44**, 591–615; (*b*) J. J. Beintema, *FEBS Lett.*, 1994, **350**, 159–163.
9 (*a*) J. Martins, D. Maes, R. Loris, H. A. M. Pepermans, L. Wyns, R. Willen and P. Verheyden, *J. Mol. Biol.*, 1996, **258**, 322–333; (*b*) R. T. Lee, H. J. Gabius and Y. C. Lee, *Glycoconjugate J.*, 1998, **15**, 649–655.
10 (*a*) C. S. Wright, *J. Mol. Biol.*, 1984, **178**, 91–104; (*b*) C. S. Wright, *J. Mol. Biol.*, 1990, **215**, 635–51; (*c*) C. S. Wright, *J. Biol. Chem.*, 1992, **267**, 14345–14352.
11 A. Rodríguez, M. Tablero, B. Barragan, P. Lara, M. Rangel, B. Arreguin, L. Possani and M. Soriano-Garcia, *J. Cryst. Growth*, 1986, **76**, 710–714.
12 A. Rodríguez-Romero, K. G. Ravichandran and M. Soriano-Garcia, *FEBS Lett.*, 1991, **291**, 307–309.
13 (*a*) J. L. Asensio, F. J. Cañada, M. Bruix, A. Rodriguez-Romero and J. Jiménez-Barbero, *Eur. J. Biochem.*, 1995, **230**, 621–633; (*b*) J. L. Asensio, F. J. Cañada, M. Bruix, C. Gonzalez, N. Khiar, A. Rodríguez-Romero and J. Jiménez-Barbero, *Glycobiology*, 1998, **8**, 569–577.
14 N. H. Andersen, B. Cao, A. Rodríguez-Romero and B. Arreguin, *Biochemistry*, 1993, **32**, 1407–1422.
15 J. L. Asensio, F. J. Cañada, H. C. Siebert, J. Laynez, A. Poveda, P. M. Nieto, U. M. Soedjanaamadja, J. J. Beintema, H. J. Gabius and J. Jiménez-Barbero, *Chem. Biol.*, 2000, **7**, 529–543.
16 (*a*) P. Sedmera, V. Prikrylová, K. Bezouška, E. Rajnochová, J. Thiem and V. Kren, *J. Carb. Chem.*, 1998, **17**, 1351–1357; (*b*) C. A. Lobry de Bruyn and W. A. van Ekenstein, *Rec. Trav. Chim. Pays-Bas.*, 1899, **18**, 147–149.
17 L. Hušáková, E. Herkommerová-Rajnochová, T. Semenuk, M. Kuzma, J. Rauvolfová, V. Prikrylová, R. Ettrich, O. Plíhal, K. Bezouška and V. Kren, *Adv. Synth. Catal.*, 2003, **345**, 735–742.
18 V. Kren and J. Thiem, *Chem. Soc. Rev.*, 1997, **26**, 463–474.
19 S. Singh, J. Packwood, C. J. Samuel, P. Critchley and D. H. G. Crout, *Carbohydr. Res.*, 1995, **279**, 293–305.

20 N. L. Allinger, Y. H. Yuh and J. H. Lii, *J. Am. Chem. Soc.*, 1989, **111**, 8551–8566.
21 F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, C. Caufield, G. Chang, T. Hendrickson and W. C. Still, *J. Comput. Chem.*, 1990, **11**, 440–467.
22 J. F. Espinosa, J. L. Asensio and M. Bruix and J. Jiménez-Barbero, *An. Quim.*, 1996, **92**, 320–324.
23 For a discussion on the application of molecular mechanics force fields to sugar molecules, see S. Perez, A. Imberty, S. Engelsen, J. Gruza, K. Mazeau, J. Jiménez-Barbero, A. Poveda, J. F. Espinosa, B. P. van Eyck, G. Johnson, A. D. French, M. Louise, C. E. Kouwijzer, P. D. J. Grootenuis, A. Bernardi, L. Raimondi, H. Senderowitz, V. Durier, G. Vergoten and K. Rasmussen, *Carbohydr. Res.*, 1998, **314**, 141–155.
24 W. C. Still, A. Tempczyk, R. C. Hawley and T. Hendrickson, *J. Am. Chem. Soc.*, 1990, **112**, 6127–6128.
25 L. M. Mikkelsen, M. J. Hernáiz, M. Martín-Pastor, T. Skrydstrup and J. Jiménez-Barbero, *J. Am. Chem. Soc.*, 2002, **124**, 14940–14951.
26 D. A. Cumming and J. P. Carver, *Biochemistry*, 1987, **26**, 6664–6676.
27 J. L. Asensio, F. J. Cañada and J. Jimenéz-Barbero, *Eur. J. Biochem.*, 1995, **233**, 618–630.
28 J. L. Asensio, A. Gracia, M. T. Murillo, A. Fernández-Mayoralas, F. J. Cañada, C. R. Johnson and J. Jimenéz-Barbero, *J. Am. Chem. Soc.*, 1999, **121**, 11318–11329.
29 (*a*) R. U. Lemieux, *Acc. Chem. Res.*, 1996, **29**, 373–380; (*b*) J. P. Carver, S. W. Michnick, A. Imberty and D. A. Cumming, *Ciba Found. Symp.*, 1991, **158**, 6–26; (*c*) M. S. Searle and D. H. Williams, *J. Am. Chem. Soc.*, 1992, **114**, 10690–10697; (*d*) H. J. Gabius, *Pharm. Res.*, 1998, **15**, 23–30.
30 (*a*) K. A. Kronis and J. P. Carver, *Biochemistry*, 1985, **24**, 834–840; (*b*) K. A. Kronis and J. P. Carver, *Biochemistry*, 1985, **24**, 826–833.
31 (*a*) D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. DeBolt, D. Ferguson, G. Siebal and P. A. Kollmann, *Comput. Phys. Commun.*, 1995, **91**, 1–41; (*b*) D. A. Pearlman and P. A. Kollmann, *J. Mol. Biol.*, 1991, **220**, 457–479.
32 D. Uhrín and P. N. Barlow, *J. Magn. Reson.*, 1997, **126**, 248–255.
33 C. E. A. F. Schafmeister, W. S. Ross and V. Romanovski, *LEAP*, 1995, University of California, San Francisco.
34 K. Bezouška, G. Vlahas, O. Horváth, G. Jinochová, A. Fišerová, R. Giorda, W. H. Chambers, T. Feizi and M. Pospíšil, *J. Biol. Chem.*, 1994, **269**, 16945–16952.
35 (*a*) P. Krist, E. Herkommerová-Rajnochová, J. Rauvolfová, T. Semenuk, P. Vavrušková, J. Pavlícek, K. Bezouška, L. Petruš and V. Kren, *Biochem. Biophys. Res. Commun.*, 2001, **287**(1), 11–20; (*b*) K. Bezouška, J. Sklenár, J. Dvoráková, V. Havlícek, M. Pospíšil, J. Thiem and V. Kren, *Biochem. Biophys. Res. Commun.*, 1997, **238**, 149–153.
36 Z. Hunková, V. Kren, M. Ščigelová, L. Weignerová, O. Scheel and J. Thiem, *Biotechnol. Lett.*, 1996, **18**, 725–730.
37 L. Weignerová, P. Vavrušková, A. Pišvejcová, J. Thiem and V. Kren, *Carbohydr. Res.*, 2003, **338**, 1003–1008.
38 Bezouška and V. Kren, DDBJ/EMBL/GeneBank databases under the accession No. AY091636.
39 K. Bock and Ch. Pedersen, *J. Chem. Soc., Perkin Trans. 2*, 1974, 293–297.