Structural basis for chitin recognition by defense proteins: GlcNAc residues are bound in a multivalent fashion by extended binding sites in hevein domains

Juan L Asensio[1], Francisco J Cañada[1], Hans-Christian Siebert[2], José Laynez[3], Ana Poveda[4], Pedro M Nieto[5], UM Soedjanaamadja[6], Hans-Joachim Gabius[2] and Jesús Jiménez-Barbero[1]

Background: Many plants respond to pathogenic attack by producing defense proteins that are capable of reversible binding to chitin, a polysaccharide present in the cell wall of fungi and the exoskeleton of insects. Most of these chitin-binding proteins include a common structural motif of 30 to 43 residues organized around a conserved four-disulfide core, known as the 'hevein domain' or 'chitin-binding' motif. Although a number of structural and thermodynamic studies on hevein-type domains have been reported, these studies do not clarify how chitin recognition is achieved.

Results: The specific interaction of hevein with several $(GlcNAc)_n$ oligomers has been studied using nuclear magnetic resonance (NMR), analytical ultracentrifugation and isothermal titration microcalorimetry (ITC). The data demonstrate that hevein binds $(GlcNAc)_{2-4}$ in 1:1 stoichiometry with millimolar affinity. In contrast, for $(GlcNAc)_5$, a significant increase in binding affinity is observed. Analytical ultracentrifugation studies on the hevein–$(GlcNAc)_{5,8}$ interaction allowed detection of protein–carbohydrate complexes with a ratio of 2:1 in solution. NMR structural studies on the hevein–$(GlcNAc)_5$ complex showed the existence of an extended binding site with at least five GlcNAc units directly involved in protein–sugar contacts.

Conclusions: The first detailed structural model for the hevein–chitin complex is presented on the basis of the analysis of NMR data. The resulting model, in combination with ITC and analytical ultracentrifugation data, conclusively shows that recognition of chitin by hevein domains is a dynamic process, which is not exclusively restricted to the binding of the nonreducing end of the polymer as previously thought. This allows chitin to bind with high affinity to a variable number of protein molecules, depending on the polysaccharide chain length. The biological process is multivalent.

Addresses: [1]Instituto de Química Orgánica General, CSIC, 28006 Madrid, Spain. [2]Institut für Physiologische Chemie, Tierärztliche Fakultät, Ludwig-Maximilians-Universität, D-80539 München, Germany. [3]Instituto Química Física Rocasolano, CSIC, 28006 Madrid, Spain. [4]SIDI-Universidad, Autónoma de Madrid, Cantoblanco, Madrid, Spain. [5]Instituto Investigaciones Químicas, CSIC, Isla de la Cartuja, 41092 Sevilla, Spain. [6]Laboratorium Biokimia, Padjadjaran University, 40191 Bandung, Indonesia.

Correspondence: Jesús Jiménez-Barbero
E-mail: iqoj101@iqog.csic.es

Key words: analytical ultracentrifugation, chitin recognition, hevein domains, microcalorimetry, NMR, protein–carbohydrate interactions

Received: **17 February 2000**
Revisions requested: **23 March 2000**
Revisions received: **10 April 2000**
Accepted:**13 April 2000**

Published: **26 June 2000**

Chemistry & Biology 2000, **7**:529–543

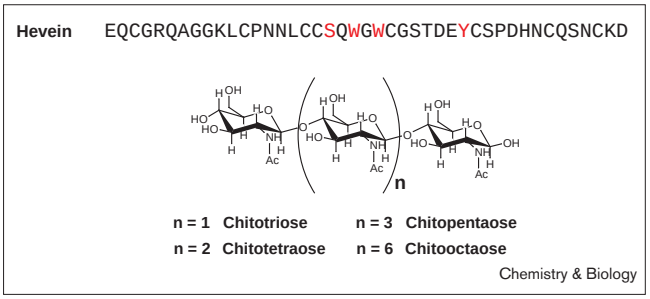
1074-5521/00/$ – see front matter

Introduction

Carbohydrates are important in energy storage and as constituents of the structural framework of cells and tissues. Owing to their extraordinary capacity to encode information stereochemically, saccharides take part in a wide variety of recognition processes of biological significance. Carbohydrate recognition by proteins has been shown to be involved in viral and microbial infection, plant defense, inflammatory responses, innate immunity, fertilization, tumor spread and growth regulation [1–6]. 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 keen interest [7–10]. Detailed information on the three-dimensional (3D) structure of protein–carbohydrate complexes has frequently been obtained from X-ray [11–15] and modeling [16], but the commonly high molecular weight of lectins has prevented their direct study using nuclear magnetic resonance (NMR). In favorable cases, however, NMR may provide information about the structural features of protein–carbohydrate complexes in solution [17–21].

Among the biological processes mentioned before, many plants respond to pathogenic attack by producing defense proteins [22] that are able to bind reversibly to chitin, a β(1→4) linked *N*-acetyl glucosamine (GlcNAc) polysaccharide (Figure 1). This natural biopolymer is a key structural component of the cell wall of fungi and the exoskeleton of invertebrates, such as insects and nematodes. Most of these defense proteins include a common structural motif of 30 to 43 residues that are rich in glycines and cysteines in highly conserved positions which are organized around a

Figure 1



Representation of hevein and the chitin fragments used in this study. Residues relevant for sugar binding are highlighted in red.

four-disulfide core, usually known as the hevein domain or chitin-binding motif (CBD). The name stems from hevein, a small protein of 43 amino acids (Figure 1) that is found in laticifers of the rubber tree (*Hevea brasiliensis*). The hevein domain is also present in several other lectins such as pseudohevein, *Urtica dioica* agglutinin (UDA), wheat germ agglutinin (WGA) and Ac-AMP antimicrobial peptides. The same CBD can also be found in enzymes with antifungal activity, such as 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 CBDs, like hevein itself or Ac-AMP peptides, have been shown to have remarkable antifungal and antinutrient activity in insects [22], even though they do not have any known enzymatic activity. Although the molecular basis for this biological activity remains unclear, the binding of the polypeptide to chitin should be the first step in this process.

Several structural and thermodynamic studies on CBDs have been reported. NMR titration [23–27] and isothermal titration calorimetry (ITC) [28,29] experiments showed that these proteins bind to $(GlcNAc)_n$ oligomers with millimolar affinity, and that the binding strength increases one order of magnitude per GlcNAc unit for n = 1–3. This experimental observation led to the proposal of a three-subsite model for chitin recognition. However, both X-ray and NMR studies have allowed the characterization of only two of these subsites at the structural level. Thus, according to NMR [23–25], laser photo-CIDNP [30] and X-ray [31–33] studies on both hevein and WGA, the aromatic residues at relative positions 21, 23 and 30 in hevein domains are essential in carbohydrate binding, because they stabilize the complexes through CH-π stacking interactions and van der Waals contacts. Hydrogen bonding between a serine at position 19 and the acetamide moiety of one GlcNAc residue is also operative. A second hydrogen bond between a tyrosine at position 30 and OH3 of the same GlcNAc unit would also take place, according to X-ray studies on WGA–sugar complexes [31–33]. No direct experimental evidence for this interaction in solution has been reported so far.

All the structural information available for complexes between hevein domains and sugars [23,24,31–33] has been obtained using very short $(GlcNAc)_n$ oligomers, with n = 1 or 2. According to these data, the strongest interactions between hevein domains and β(1→4) linked oligosaccharides comes from the terminal nonreducing sugar residue in subsite +1 [31]. Modeling experiments carried out on WGA suggested that binding of interior GlcNAc units in the polymer at subsite +1 could be forbidden, because of steric hindrance when the oligosaccharide is extended at position C4 [31–33]. On these grounds, chitin recognition by hevein-type domains would only be allowed at the terminal nonreducing ends of the polymer. However, this hypothesis has difficulty explaining the efficient recognition of the chitin chain, and thus the biological action of these proteins. Although several NMR studies have analyzed $(GlcNAc)_3$ binding to hevein and AcAMP-2 [24,25], a proper description of the sugar location in the binding site has not been provided. This is because of the rather low stability of the complexes (usually high ligand:protein ratios have to be used), and the extreme broadening of the sugar signals due to the free–bound exchange process.

In general, the use of short GlcNAc oligomers makes the analysis of the protein–ligand interaction easier, and provides a simple model for the study of chitin binding. Nevertheless, many biological interactions may be multivalent in nature [34–37], and additional protein–carbohydrate interactions could potentially be generated for longer chitin fragments, thus significantly affecting the thermodynamic balance of the recognition process and/or the 3D structure of the protein. On this basis, longer $(GlcNAc)_n$ oligomers with $n>2$ would constitute a more appropriate model of chitin and should provide a deeper understanding of chitin recognition and thus, of plant defense mechanisms.

Here we have made a thermodynamic ITC and structural (NMR) study of the interaction of hevein with several oligomers of $(GlcNAc)_n$, with n = 1–5 (N,N′,N″,N‴,N⁗-pentacetyl-chitopentaose). In addition, the average molecular weight of different hevein complexes with $(GlcNAc)_n$ oligomers, up to n = 8, has been deduced by analytical ultracentrifugation. Finally, the detailed structural study of the hevein–$(GlcNAc)_5$ complex has been carried out using NMR, with a ^{13}C-labeled pentasaccharide. Using all these data, we present a general and structurally detailed model for chitin recognition by hevein domains.

Results and discussion

Ligand-binding studies: ITC and sedimentation equilibrium experiments

The thermodynamic parameters K_a, ΔH, and ΔS for the interaction of hevein with $(GlcNAc)_n$ oligomers (up to

$n = 5$) were measured using ITC at 25°C. The derived values (Table 1, Figure S1 in the Supplementary material section) for the di- and tri-saccharide are in agreement with our previous NMR data [23,24] and show that the affinity of the lectin for these ligands is in the millimolar range, the recognition process is enthalpically driven, and that entropy opposes binding. Our enthalpy values are slightly larger than those previously obtained using microcalorimetry by Garcia-Hernandez *et al.* [29]. In any case, according to both ITC and NMR, the affinity of the protein for the ligand increases one order of magnitude per GlcNAc residue between $n = 1$ and $n = 3$. In terms of ΔG, this increase is the consequence of a more favorable enthalpic balance (Table 1) for the longer $(GlcNAc)_n$ oligomers. A similar behavior has been observed in other hevein domains, leading to the proposal of a three-subsite model for chitin recognition [38]. Previous X-ray [31–33] and NMR studies [23,24] on these domains have clearly identified two subsites at the structural level, defined by aromatic rings (Trp or Tyr) at positions 23 (referred as subsite +1) and 21 (subsite +2). subsite +1 also involves Ser19 and Tyr30 and provides the key interactions for the binding of $(GlcNAc)_n$ oligomers (Figure S2a).

For tetrasaccharide binding ($n = 4$), a further increase in ΔH of about 1 kcal/mol is observed in comparison to the trisaccharide ($n = 3$). In contrast with the observed behavior for shorter oligomers, this favorable $\Delta\Delta H$ is almost completely counterbalanced by the entropic contribution, thus leading to a negligible increase in the association constant, K_a. It has usually been assumed that hevein domains exclusively interact with the nonreducing end of chitin oligomers [31–33]; however, a model of the complex based on this assumption (Figure S2b) shows that the reducing end of the tetrasaccharide is completely exposed to the solvent and does not make any contact with the protein. Therefore, the observed $\Delta\Delta H$ value cannot be satisfactorily explained. Moreover, the ITC data indicate a sharp increase in the K_a value when pentasaccharide binding is monitored. The K_a increased from 11,000 M^{-1} (tetra) to more than 450,000 M^{-1} (penta). In this case, and opposite to the observed behavior for $n = 1$–4, it was not possible to get a perfect fit of the experimental ITC curves by assuming a pure 1:1 stoichiometry (Figure S1). This result strongly suggests the existence of higher order complexes in solution for pentasaccharide association.

In order to test this hypothesis, the average molecular weights of several hevein–carbohydrate complexes were measured by analytical ultracentrifugation at different protein:ligand ratios. A representation of the dependence of the average molecular weight on the protein:ligand ratio for the different ligands is shown in Figure 2 (see also Table S1). For hevein binding to $(GlcNAc)_3$, the ultracentrifugation experiments conclusively demonstrate the exclusive existence of 1:1 complexes in solution (average molecular weight below 5300 Da under all experimental condition tested). In contrast, for $(GlcNAc)_5$ binding, significantly larger molecular weights, incompatible with an unique 1:1 stoichiometry, were obtained (up to 7600 Da, depending on the protein:carbohydrate ratio and the total protein concentration). The observed dependence of the average molecular weight of the complexes on the protein–ligand ratio used strongly suggests the presence of protein–ligand complexes in solution with 2:1 stoichiometry, thus reflecting the existence of a mixture of 1:1 and 2:1 protein–sugar complexes. Probably, for chitin fragments up to $n = 4$, the carbohydrate length is too short to allow the binding of two hevein moieties. In contrast, for $(GlcNAc)_5$, the carbohydrate chain may comprise more than one binding site, thus allowing the interaction with two hevein molecules. According to this reasoning, further increase in the average molecular weight would be expected for longer $(GlcNAc)_n$ oligomers. Ultracentrifugation experiments carried out with $(GlcNAc)_8$, also at different protein:ligand ratios (Figure 2 and Table S1), indicated the presence of larger molecular weights (up to 10,100, depending on the protein:carbohydrate ratio). This result is consistent with the major presence of 2:1 complexes in solution, although the existence, to some extent, of 3:1 complexes cannot be excluded. The nature, at least in part, of multivalent interactions between hevein and long $(GlcNAc)_n$ oligomers is thus demonstrated.

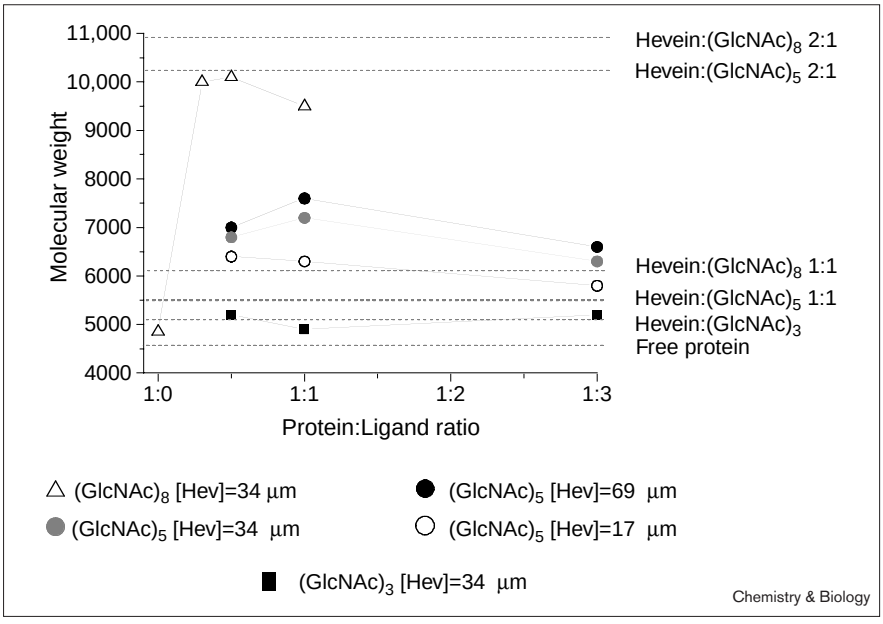
Table 1

Association constants (K_a), and thermodynamic parameters ΔH and ΔS for the binding of $(GlcNAc)_{2-5}$ to hevein, obtained using ITC and NMR.

Ligand	K_a (M^{-1})	ΔG (kcal/mol)	ΔH (kcal/mol)	ΔS (cal/mol K)
$(GlcNAc)_2$ ITC	616	−3.8	−6.3	−8.4
$(GlcNAc)_3$ ITC	8525	−5.4	−8.3	−9.9
$(GlcNAc)_4$ ITC	10,850	−5.5	−9.5	−13.4
$(GlcNAc)_5$ ITC	474,000	−7.8	−9.6	−6.3
$(GlcNAc)_5$ ITC	587,000	−7.9	−10.1	−7.5
(*Inverse titration*)				
GlcNAc NMR	30	−2.0	–	–
LacNAc NMR*	115	−2.8	–	–
$(GlcNAc)_2$ NMR	602	−3.8	−7.5	−12.4
$(GlcNAc)_3$ NMR	11,558	−5.5	−8.7	−10.7

*The NMR-derived K_a value for LacNAc binding to hevein is also shown. K_a values and van′t Hoff ΔH and ΔS values for $(GlcNAc)_{2-3}$, as derived by NMR are shown for comparison. For the ITC titrations, the $(GlcNAc)_{2-4}$ solutions were added to the protein solution. For $(GlcNAc)_5$, additional titrations using the inverse protocol were carried out. In these experiments, the protein solution was added to the ligand solution in the cell. All experiments were repeated at least three times and the average is shown.

Figure 2



Representation of the average molecular weight corresponding to the different protein–sugar complexes, as obtained by analytical ultracentrifugation. Different protein concentrations and protein:ligand ratios were used.

The assumption that hevein domains mainly interact with the nonreducing end of chitin oligomers is hardly compatible with the experimental observation of 2:1 protein–carbohydrate complexes in solution. Furthermore, no obvious structural explanation based on this hypothesis can be given for the $\Delta\Delta H$ increase observed for tetrasaccharide versus trisaccharide binding. Similarly, the sharp increase in affinity of the protein for $(GlcNAc)_5$ cannot be easily explained assuming the exclusive interaction of the nonreducing unit with subsite +1. Therefore, in order to verify whether a GlcNAc residue was indeed necessary at the nonreducing terminal end to achieve binding, we titrated hevein with *N*-acetyllactosamine (LacNAc) and monitored the chemical-shift changes by NMR (Figure S3). The results unequivocally demonstrate that LacNAc, with galactose (Gal) and not GlcNAc at the nonreducing end, was indeed bound with fourfold affinity ($K_a = 120\ M^{-1}$) relative to the monosaccharide ($K_a = 30\ M^{-1}$ [23]). The additional stabilization by 3.5 kJ/mol infers two conclusions: hevein domains may bind GlcNAc-containing oligosaccharides using residues other than the nonreducing end, and assuming that the GlcNAc residue is bound at subsite +1, other subsites (subsites –1,–2 or even –3) could be operative (see below).

The data for LacNAc binding also have implications in the interactions between $(GlcNAc)_n$ and hevein domains. The thermodynamic parameters for the binding of long $(GlcNAc)_n$ fragments could reflect the existence of a mixture of 1:1 complexes in solution, with either the nonreducing end or different GlcNAc units bound at subsite +1. Moreover, for the pentasaccharide, the length of the ligand allows the binding of two hevein molecules, thus producing a mixture of several 1:1 and 2:1 protein–carbohydrate complexes.

Structural basis for chitin recognition

The protein

As a first step, we monitored NMR chemical-shift variations induced in the protein as a result of complexation. In contrast with the behavior reported for the binding of hevein to shorter oligomers, for which the exchange rate is fast on the chemical-shift timescale, in the case of $(GlcNAc)_5$ the exchange rate between the hevein free and bound states at 25°C is in the intermediate–slow regime (Figure S4), thus making impossible the derivation of K_a values from NMR titrations. This experimental observation is consistent with the higher affinity of the protein for this extended ligand. Interestingly, two different signals are observed for the aromatic protons Hε1 and Hε2 of residue Y30 at 5°C (Figure S5). Unambiguous exchange-mediated cross-peaks between both Y30 proton signals are observed in ROESY (Figure S5) spectra at 5°C with a protein:$(GlcNAc)_5$ ratio of 1:2. At higher temperatures, both singlets coalesce and provide one average signal at 25°C. In contrast, only one average signal is observed for both protons in the experiments corresponding to hevein in the free state at 5°C. The fact that the rotation of Y30 is slow on the chemical-shift timescale, but only when complexed to the pentasaccharide, reflects the freezing of the Y30 sidechain as a consequence of its interaction with the carbohydrate. The restriction of flexibility of the hevein sidechains owing to sugar binding has probably a significant effect on the entropic balance of the recognition process. In

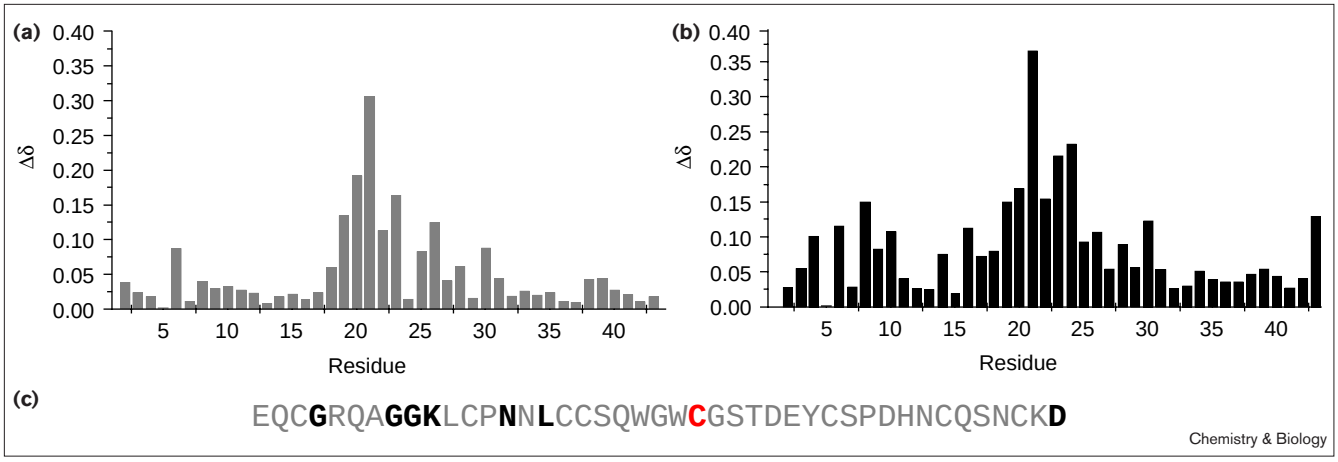
fact, this effect has been considered by several authors to be the main origin of the entropy–enthalpy compensation phenomenon [15,39–41], usually observed in protein–sugar interactions. Additionally, exchange cross-peaks were also observed for protons at the indol ring of W21, a residue that is also directly involved in sugar recognition [24]. This fact indicates the existence of two orientations for this aromatic system in the complexes (see below).

The assignment of ^{1}H-NMR spectrum of hevein in the complex with $(GlcNAc)_5$ was based on our previous data for free hevein and hevein bound to $(GlcNAc)_2$ [23,24]. Despite the evident broadening of some signals, the almost complete assignment of the spectrum was successfully accomplished at 5°C, using well-established protocols with nuclear Overhauser effect spectroscopy (NOESY)/total correlation spectroscopy (TOCSY) experiments at different protein:ligand ratios, between 1:2 and 1:4, for which the 1:1 complex is predominant (see the Materials and methods). The ^{1}H-NMR assignment and the structural information derived from the NMR experiments therefore correspond basically to protein–ligand complexes with 1:1 stoichiometry. Analysis of different NOESY spectra allowed the unambiguous assignment of 350 NOEs between protein proton pairs, already observed [24] in hevein complexes with $(GlcNAc)_2$ and $(GlcNAc)_3$. This result rules out the existence of major differences for the protein structure among the three $(GlcNAc)_2$, $(GlcNAc)_3$, and $(GlcNAc)_5$ complexes.

In principle, an estimation of the protein regions involved in contacts with the ligand can be obtained by analyzing the chemical-shift changes induced by the association process, as previously applied to the study of the hevein–chitobiose complex, which allowed the binding site to be diagnosed [23]. For $(GlcNAc)_5$, additional protein–carbohydrate interactions might be expected, as the thermodynamic balance of the recognition process is far different from that for $(GlcNAc)_{2–3}$. Thus, we compared the induced $\Delta\delta$ (δ_{free}–δ_{bound}) for the hevein–$(GlcNAc)_5$ complex with those for the hevein–$(GlcNAc)_3$ complex to characterize any additional amino-acid residue involved in sugar recognition (Figure 3). In both cases, the most affected residues are located between S19 and Y30. This region includes all the amino acids involved in sugar binding according to previous studies [24]. Nevertheless, clear differences in the induced $\Delta\delta$ can be observed between both complexes. Thus, the NH of C24 and the loop region between K10 and L16 give rise to large $\Delta\delta$ differences for hevein–$(GlcNAc)_5$. In contrast, the corresponding $\Delta\delta$ values for hevein–$(GlcNAc)_3$ are negligible. This experimental observation suggests that additional protein regions are involved in sugar binding, but only for long GlcNAc oligomers.

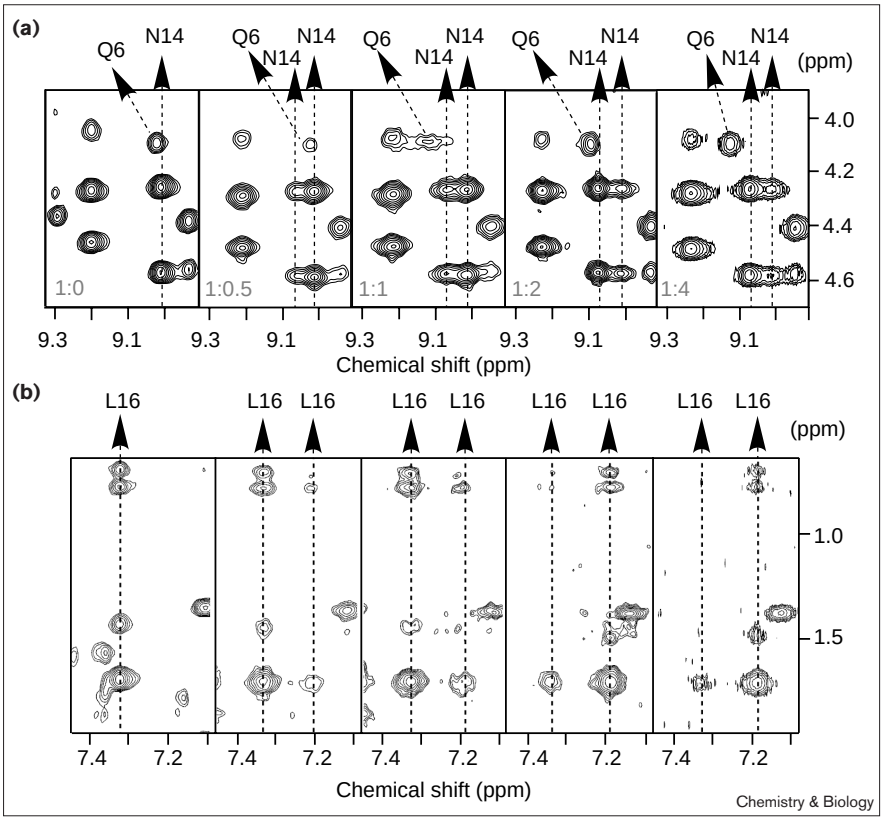
A closer inspection of the protein spectra provides additional evidence for the participation of the P13–L16 loop in carbohydrate interaction and for the existence of more than one 1:1 protein–sugar complex in solution. Indeed, two sets of slow exchanging signals can be observed for the NHs of residues N14, N15 and L16 (Figure 4). One of the sets corresponds to the free state of the protein and, also, to those complexes for which the NHs of N14, N15 and L16 are not affected with respect to the free state (complexes type I, Figure 5). The second set corresponds to those complexes for which these residues are perturbed

Figure 3



Chemical shift changes for ligand binding to hevein. **(a)** Representation of the change in chemical shift ($\Delta\delta$; δbound–δfree) observed for the hevein backbone protons in the complex with $(GlcNAc)_3$. Only the largest $\Delta\delta$ (HN or Hα) is represented for every residue. For clarity, all $\Delta\delta$ values are represented as positive. **(b)** $\Delta\delta$ (δbound–δfree) observed for hevein backbone protons in the complex with $(GlcNAc)_5$. **(c)** Hevein sequence highlighting in black the residues whose behavior clearly differs in both complexes. The residue that presents the largest difference in $\Delta\delta$ between both complexes is highlighted in red.

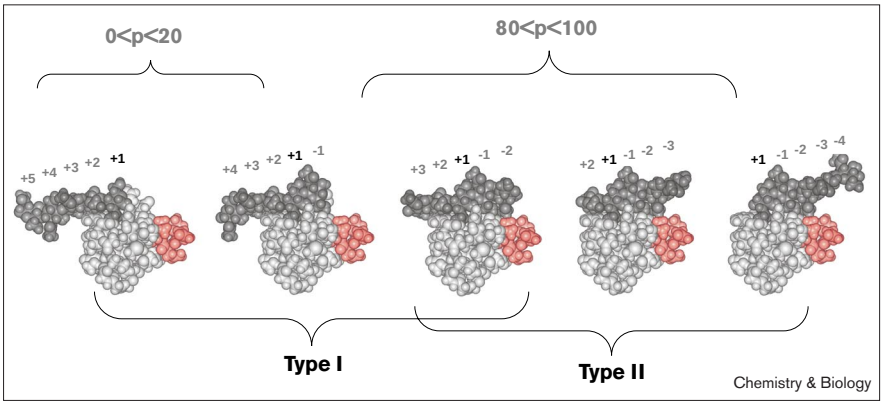
Figure 4



NOESY spectra obtained at 5°C for different hevein:$(GlcNAc)_5$ ratios. The effect of ligand binding on the HN signals of **(a)** N14 and **(b)** L16 is shown. Two sets of signals can be observed at a protein:ligand ratio of 1:4 for the two HN protons. This fact indicates the existence of two kind of complexes in solution. The L16/N14 region remains unaffected by the binding process in one, but is perturbed in the other.

by the binding process (complexes type II). It is important to notice that L16 chemical shift is only significantly affected in the presence of the pentasaccharide, and not the trisaccharide. NOESY experiments carried out at different protein:$(GlcNAc)_5$ ratios reveal that the relative intensities of both signals is only dependent on the protein:$(GlcNAc)_5$ ratio, up to 1:2. No further change is observed for higher molar fractions of $(GlcNAc)_5$ (Figure 4). This fact reveals that even when the protein is saturated with sugar, there is still a certain amount of complexes in solution for which the N14–L16 region remains unaffected (complexes type I). The fraction of these complex(es) was estimated from the intensity ratio between both sets of NH signals, as measured in NOESY spectra acquired for a hevein:$(GlcNAc)_5$ ratio of 1:4. The proportion of complexes type I is 20%, and that of type II

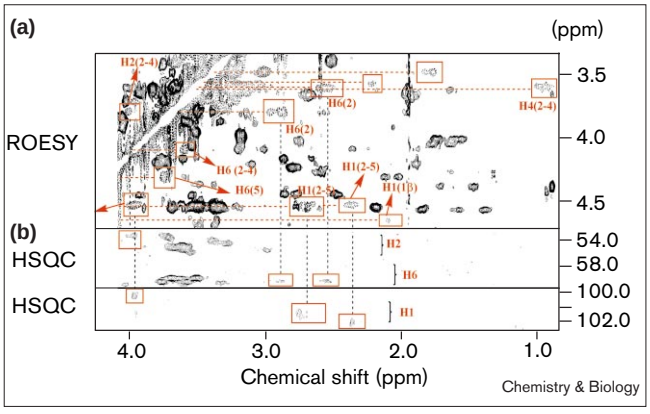
Figure 5



CPK representation of the models generated for the hevein–$(GlcNAc)_5$ complexes, assuming the binding of every single GlcNAc unit (from 1 to 5) at subsite +1. Complexes in which the sugar lies far from the L16/N14 region are labeled 'type I'. Complexes in which the ligand lies close to that region are labeled 'type II'. The relative amounts of both types of complexes can be quantified by NMR (see Figure 4). The loop region 13–16 is highlighted in red.

is 80%. In order to interpret this experimental result, we modeled five different hevein–$(GlcNAc)_5$ complexes by docking every single GlcNAc unit at subsite +1 of the protein, using the NMR structure of hevein [24]. Figure 5 shows the models (a–e) generated for these five possible 1:1 complexes. It can be observed that, in complexes a and b, all the GlcNAc units are located too far from the P13–L16 loop. In contrast, a GlcNAc moiety at position –2 (complexes c–e) and/or –3 (complexes d and e) could be the cause of the observed effect on P13–L16 chemical shifts. The modeling in combination with the NMR data indicates that the population of complexes a+b is smaller than 20%, whereas that of complexes c+d+e is higher than 80%. The maximum fraction of complexes a+b is with certainty, not higher than 20%, which indicates that the recognition of the nonreducing end constitutes a minor event for long $(GlcNAc)_n$ oligomers. Although no quantitative conclusions can be extracted from these NMR experiments, it seems reasonable that the recognition at position +1, of either the reducing or the nonreducing end sugar moieties, in fact represents an insignificant proportion of the obtained 80–100% and 0–20% fractions, respectively.

Figure 6



ROESY spectrum **(a)**, corresponding to the hevein:$(GlcNAc)_5$ complex (1:2 molar ratio) at 5°C, showing free–bound exchange peaks for the sugar signals. Huge changes in chemical shift for the bound ligand can be observed in some cases (boxed regions). These exchange peaks were used to assign the pentasaccharide in the complexes with hevein. As the changes in ^{13}C chemical shift are usually small, HSQC spectra **(b)** were used to resolve ambiguities.

From the biological point of view, these experimental results are in agreement with the fact that hevein is a chitin-binding protein — its targets being oligomeric structures. It would indeed be difficult to explain its action as defense protein if it could only bind terminal nonreducing ends of GlcNAc-containing oligosaccharides. We have shown that the extension of the polysaccharide from the GlcNAc unit at position +1 in both directions generates additional stabilizing protein–carbohydrate interactions, thus allowing a more efficient binding of hevein domains to chitin.

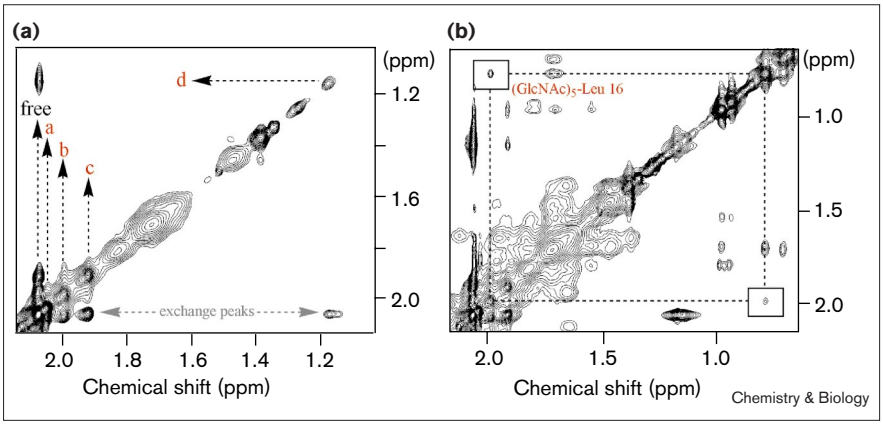
The sugar

In order to derive the 3D structure of the complex in solution, we have to analyse the sugar resonances and the sugar–protein NOEs. As a first step, the ^1H, ^{13}C-NMR assignments of the free pentasaccharide were obtained (Figure S6 and Table S2) from the analysis of different HSQC, HSQC-TOCSY and HSQC-NOESY experiments (see the Materials and methods). Once the free sugar was analyzed, the interaction between the ^{13}C-labeled $(GlcNAc)_5$ and hevein was explored. The key experiments were performed at low temperature. At 5°C, free–bound exchange-mediated cross-peaks between saccharide proton pairs were detected in both NOESY and ROESY spectra (Figure 6). Remarkable changes of the chemical shift of some sugar protons upon complexation were observed, under these slow exchange conditions. The participation of several aromatic residues in the carbohydrate recognition site accounts for this fact. The minor effect of the association on the ^{13}C chemical shifts allowed to directly assign the different carbon types. As it has been previously indicated, the detailed comparison of several NOESY/ROESY spectra revealed the existence of at least two 1:1 complexes in solution, in slow exchange on the chemical-shift timescale. For example, Figure 7 shows the acetamide methyl region as obtained in a 150 ms NOESY spectrum, at 5°C. Five different singlets between δ 1.1–2.1 ppm can be observed for the acetamide methyl groups of $(GlcNAc)_5$. Those that correspond to the free ligand (which was in excess in all these experiments) are degenerated and appear as an unique singlet at 2.05 ppm. The same chemical shift would be expected for the methyls whose chemical shifts remain unaffected by the association process (Figure 7). Four other signals in slow exchange (Figure 7a–d) can be observed to be upfield-shifted, and are assigned to bound methyl groups of $(GlcNAc)_5$. The presence of exchange cross-peaks in NOESY/ROESY spectra between two of the bound methyl signals conclusively proves the existence of a mixture of different 1:1 complexes. Therefore, each methyl signal at a given chemical shift (Figure 7a–d) cannot be assigned to any particular methyl group in $(GlcNAc)_5$ (see Materials and methods). Instead, they represent the interaction of those methyl groups with a particular subsite within the complexes. The same reasoning can be applied to the rest of the sugar signals in the bound state. The protocol for the assignment of the bound sugar signals under these conditions is described in the Materials and methods section.

The complex

The assigned protein–protein and protein–sugar NOEs were used to generate a first set of 3D models for the hevein–chitin complex. A simulated annealing protocol was employed, using the NMR structure of the protein in the hevein–methyl-β-chitobioside complex, as the starting geometry [24]. The models obtained were used to calculate

Figure 7



Evidence for exchange among different ligand–protein complexes. **(a)** NOESY spectrum corresponding to a hevein:$(GlcNAc)_5$ molar ratio of 1:4 at 5°C. The acetamide methyl region is shown. The methyl signals corresponding to the free pentasaccharide are indicated. Four additional methyl signals (labeled as i–iv) corresponding to bound sugar are present. Free–bound exchange peaks can be observed in both NOESY and ROESY spectra. Exchange peaks can be observed between the bound states (below the threshold level in the spectra, these peaks are not shown for clarity), which further demonstrates the existence of a mixture of complexes in solution. **(b)** A more extended region of the same spectrum showing the protein–sugar NOEs between the methyl signal labeled as (ii) and the sidechain of L16.

the expected Δδ for every particular sugar proton that would be induced as a result of binding to the protein. The detailed comparison of expected versus experimental Δδ values confirmed all the sugar assignments carried out in the first stage. In addition, it allowed the specific subsite assignment (see the Materials and methods section) of several new sugar signals which, in turn, allowed the detection of new protein–sugar NOEs. All the protein–protein and 22 protein–sugar NOES (Table 2) were then used in the calculation of a final NMR-derived model of the hevein–chitin complex. We obtained a set of 16 structures (Figure 8, Table S3) characterized by low energy values (AMBER forcefield [54] energies –362 to –410 kcal/mol) and free of violations larger than 0.25 Å. The pairwise r.m.s. deviation values for backbone superimposition between residues 1 and 43 and 3 and 41 were 1.0 Å and 0.8 Å, respectively. As final test for the quality of these NMR models, the induced Δδ values in the ligand caused by complexation were calculated and compared with those derived experimentally (Tables S4 and S5, Figure 9). An excellent agreement was obtained in all cases.

Table 2

Protein–sugar NOEs assigned from the inspection of several NOESY spectra at a hevein:$(GlcNAc)_5$ ratio of 1:4.

$H_{protein}$	H_{sugar}	subsite*	$H_{protein}$	H_{sugar}	subsite
Y30 OH	OH_3	+1	S19 Hβ2	CH_3	+1
Y30 OH	H_3	+1	W23 Hδ1	H2	+1
S19 OH	CH_3	+1	W23 Hε3	H5	–1
W21 Hδ1	CH_3	+1	W23 Hζ3	H5	–1
Y30 Hε1	CH_3	+1	C24 Hβ1	CH_3	–2
Y30 Hε2	CH_3	+1	C24 Hβ2	CH_3	–2
E29 Hβ1	CH_3	+1	L16 Me	CH_3	–2
E29 Hβ2	CH_3	+1	P13 Hδ1	CH_3	–2
E29 Hγ1	CH_3	+1	W23 Hζ2	H6a	+1
E29 Hγ2	CH_3	+1	W23 Hζ2	H6b	+1
S19 Hβ1	CH_3	+1	W23 Hβ2	OH3	+1

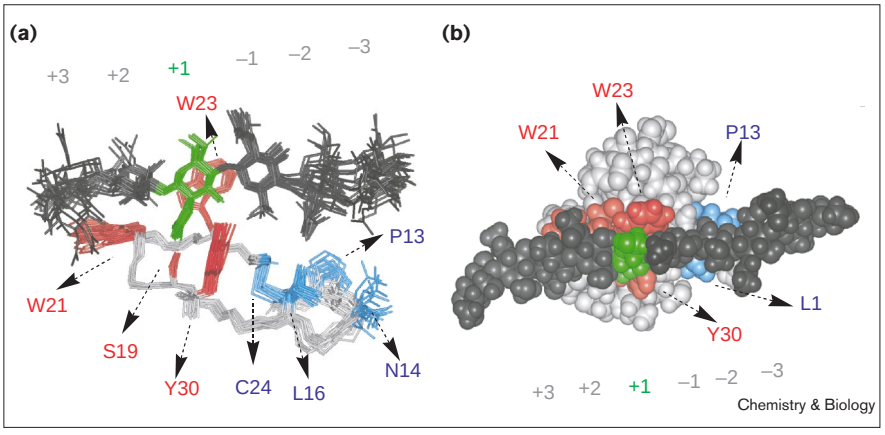
*The subsite occupied by the sugar proton involved in the protein–ligand contact is also specified.

GlcNAc at subsite +1

As previously described, most of the protein–sugar contacts are provided by subsite +1. Thus, the complex is stabilized by non-polar forces involving Trp23, Trp21 and Tyr30, and by two hydrogen bonds — one involves the sidechain of S19 and the acetamide of the GlcNAc; and the second involves the hydroxyl group of Y30 and the OH3 of the sugar moiety. The existence of both hydrogen bonds in hevein domains had been previously deduced from the X-ray analysis of different WGA–sugar complexes [31–33]. In addition, our NMR studies carried out on protein–sugar complexes with several hevein domains (i.e. hevein [24], pseudohevein and WGA-B, manuscript in preparation) conclusively proved that the first interaction also exists in solution. In fact, the OH of S19, observable for both the free and $(GlcNAc)_n$ bound states, is shifted downfield more than 1 ppm as a result of complexation. Additionally, it presents an NOE to the methyl acetamide group of the sugar at subsite +1.

No spectroscopic evidence had previously been reported for the existence of the second hydrogen bond (involving OH of Y30) in solution. For hevein–$(GlcNAc)_5$, we obtained a direct proof for this interaction in solution (Figure 10). Analysis of several NOESY spectra allowed the unambiguous assignment of the OH of Y30 and OH3

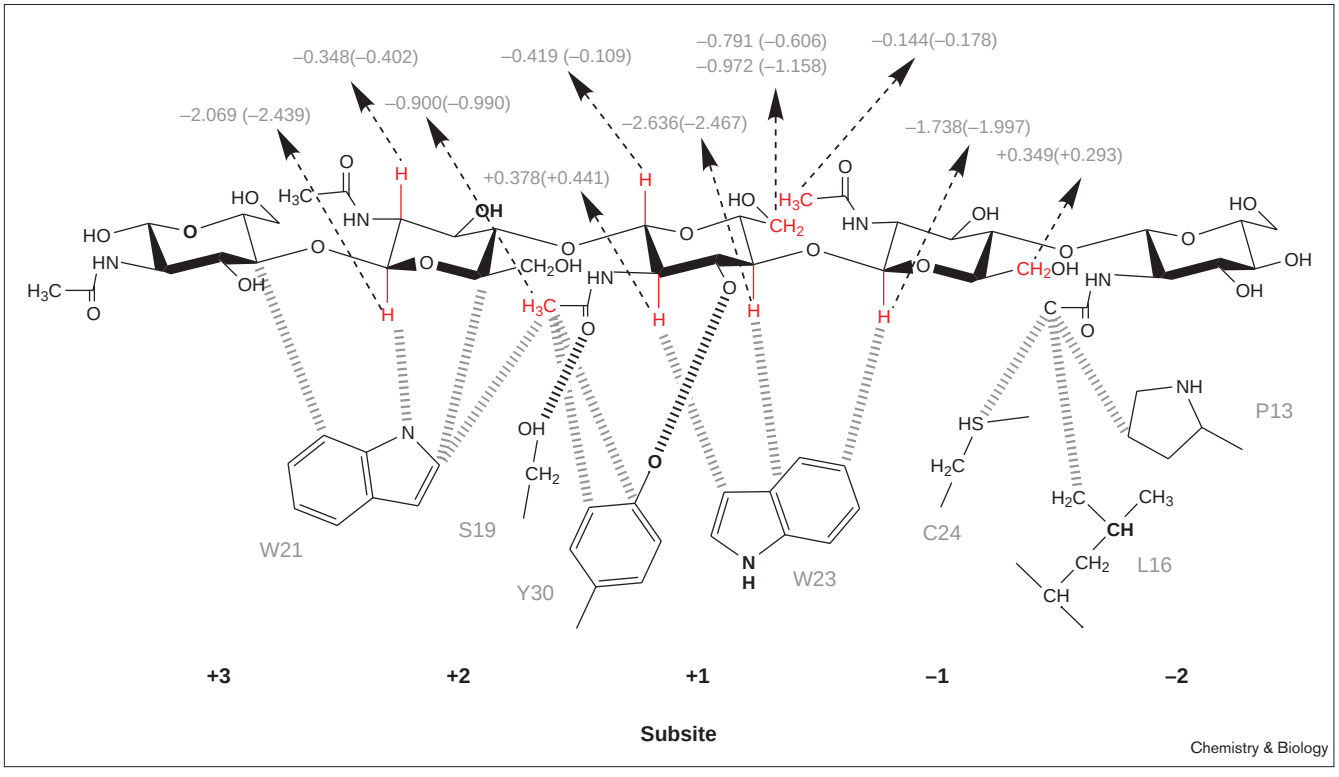
Figure 8



The hevein–chitin complex by NMR. **(a)** Ensemble of 16 structures obtained from the restrained molecular dynamics simulations. Only the binding site (residues 12–24, 30 and the sugar chain) is shown. The pairwise rms deviation values for the backbone superimposition between residues 1 and 43 and 3 and 41 are 1.0 Å and 0.8 Å, respectively. The GlcNAc unit located at subsite +1 is highlighted in green; and protein residues involved in the recognition of this GlcNAc unit are highlighted in red. Similarly, protein residues involved in recognition located in the region 12–16 are highlighted in blue. **(b)** CPK representation of the NMR-derived model of the hevein:chitin complex. The GlcNAc unit located at subsite +1 is highlighted in green. 350 protein–protein and 22 protein–sugar NOEs were used to build these models.

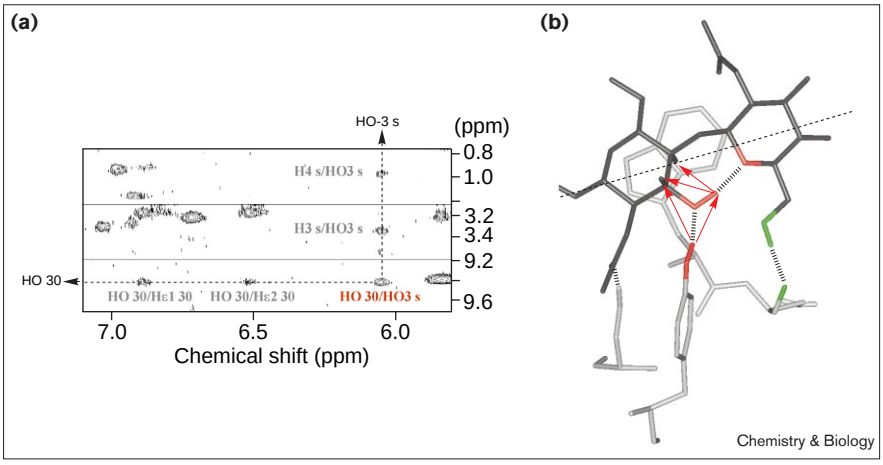
of GlcNAc at subsite +1. Moreover, protein–sugar NOEs (Figure 10) were detected between both hydroxyl groups and between the OH of Y30 and the sugar H3. The directionality of the hydrogen bond could also be determined. The OH of Y30 could only be observed in the presence of $(GlcNAc)_5$ because the free protein, even at 5°C, is in fast exchange with the solvent. This experimental observation indicates that Y30 is probably acting as hydrogen-bond donor to the HO-3 sugar acceptor in the complex. This conclusion was confirmed by the NOE data. The sugar OH3 shows NOEs with H3 and H4 of the same pyranose ring, but not with H2, indicating the existence

Figure 9



Representation of the observed protein–sugar interactions in the NMR-derived model of the hevein–chitin complex. The experimental Δδ values are represented for some of the sugar protons. The corresponding theoretical values are shown in parentheses for comparison.

Figure 10



Hydrogen bonding network from NMR. **(a)** NOESY spectra showing some of the observed cross-peaks that involve the hydroxyl proton of Y30 and the OH3 corresponding to the GlcNAc unit at position +1. The NOE between both OH groups is highlighted in red. Abbreviation: s, sugar. **(b)** Representation of the detected hydrogen bond between both OH groups. The sugar OH can be simultaneously involved in an additional hydrogen bond with the ring oxygen of the GlcNAc unit at position –1, establishing a cooperative hydrogen-bond network. An additional hydrogen bond might be established between the CH_2OH group of the GlcNAc unit at –1 and the C=O bond of C24, but no experimental evidence for this was found.

of a principal orientation for this OH, pointing towards the GlcNAc unit at subsite –1.

GlcNAc at subsite –1

According to the NMR structure, the OH of Y30 acts a hydrogen-bond donor (Figure 10), and the sugar 3-OH acts as hydrogen-bond acceptor. In addition, the sugar 3-OH could be simultaneously involved in a hydrogen bond with the pyranose ring oxygen of the GlcNAc unit at subsite –1, thus generating a cooperative hydrogen-bond network (Figure 10). The GlcNAc at –1 therefore provides additional stabilization to the complexes [42,43]. This sugar shows additional contacts to hevein. In fact, a π-CH interaction between the indol ring of W23 and the β face of this pyranose ring can be observed, reflected in the unusual chemical shift of its anomeric proton (Table S4).

GlcNAc moieties at subsites +2 and +3

The GlcNAc unit at position +2 contributes to the complex stability by a CH-π interaction between the sugar and the aromatic ring of W21. Extension of the polysaccharide towards position +3 produces additional contacts with the extended surface of the indol ring of W21. In addition, the GlcNAc unit at +3 exerts a stabilizing effect by fixing the anomeric configuration of GlcNAc at +2, in the β form, which optimizes its CH-π interaction with W21. In fact, anomerization at position +2 is expected to have a net effect on the affinity of the protein for the ligands, as shown for the recognition studies with shorter $(GlcNAc)_{2–3}$ oligomers [23,24]. Thus, the affinity of hevein for methyl-β-chitobioside (anomeric center blocked in β configuration) is three times larger than that observed for chitobiose itself (a mixture of α and β anomers). These data show the destabilizing effect of an axial anomeric OH in this protein subsite. In addition, the presence of a reducing α-GlcNAc moiety at this subsite is expected to influence the orientation of the W21 ring. This prediction was confirmed by the inspection of the ROESY spectra at 5°C. Two sets of signals, one corresponding to a minor species, were observed for this aromatic system in the bound state (Figure S5). Although it was impossible to carry out a proper structural analysis of this minor species, with its different W21 orientation, it is likely that it corresponds to the complex with a reducing α-GlcNAc unit at +2.

GlcNAc moiety at subsite –2.

There are stabilizing van der Waals interactions between the methyl acetamide group of the GlcNAc unit located at position –2 and the hydrophobic patch constituted by residues P13, L16, and C24. The participation of the loop region P13-L16 in chitin recognition has been unambiguously proved by the analysis of chemical shift changes and protein:sugar NOEs (see above).

GlcNAc at other protein subsites

The obtained structures (Figure 8) indicate that additional protein–sugar contacts might be generated by the extension of the saccharide chain towards position –3, or even –4. Although no protein–sugar NOEs in this region could be assigned, the chemical shift data are consistent with a significant degree of occupancy at subsite –3. In contrast, no additional protein–sugar interactions would be provided by the extension of the chitin chain towards position +4. According to this model structure, therefore, the minimum number of GlcNAc units required in to provide the maximum possible number of protein–sugar interactions in a 1:1 complex would be in the range of 5–7.

Sugar conformation

The NMR model also provides information on the conformation of the five GlcNAc units in the binding site. Thus, the sugar fragment that spans from +1 to +2 adopts a *syn*-type conformation, with ϕ/ψ angles of about $45 \pm 10/–10 \pm 20°$. This geometry constitutes the principal

(global minimum) conformation of chitin oligomers in solution, as determined by NMR [44]. The exclusive selection of this conformation is achieved by the particular disposition of the rings in W21 and W23. In fact, the indol plane of both residues forms an angle that matches almost perfectly the angle between the planes defined by C1–C3–C5 in both pyranoses when the global minimum *syn* conformation is considered. Therefore, both aromatic residues are only pre-organized to generate a double π/CH sugar stacking, with the major conformation in solution of the GlcNAc oligomers. In addition, in order to explain the observed sugar–sugar and protein–sugar NOEs, the conformation of the glycosidic linkages for the pentasaccharide that spans from +1 to –2 must be also located in the global minimum *syn* region, although more variability and slight deviations of ψ towards negative values could take place (ϕ/ψ about $45 \pm 20/-30 \pm 30°$). Probably, the protein reduces the entropic cost of the recognition by interacting with the most populated conformation of the chitin fragment.

Exchange process

We can envisage two possibilities for the mechanism of exchange between at least two 1:1 complexes. The first one involves the shifting of the GlcNAc rings along the subsites within the complex, that is, moving from subsite +1 to subsite –1, then to subsite –2, and so on. The second involves a relay mechanism through an in–out process. In this case, a bound GlcNAc unit at subsite +1 would go into the free state and then come back into the protein, now occupying other subsite, for example, +2 or –1. Two reasons point to the occurrence of this second possibility, illustrated in Figure S7. First, the exchange cross-peaks are much stronger for the free–bound exchange than for the bound–bound exchange. Second, for the conformation of the bound pentasaccharide, an alternation of the bulky acetamide groups take place; therefore, the shifting of the protein along the polysaccharide chain would not directly provide a good matching of the vicinal GlcNAc unit to a given subsite. Moreover, important steric conflicts of the acetamide group with the lateral chains of residues 21, 23 and 30 would take place for the shifting process, posing a high energy barrier for this mechanism.

Higher order complexes

Finally, as shown by the ultracentrifugation data, significant amounts of higher order protein–carbohydrate complexes are present in solution for protein:penta- or octa-saccharide ratios of 1:1 or larger. No structural information about these complexes could be obtained from NMR, because of the large broadening of the protein signals under all the tested experimental conditions. No higher order complexes are detected for the interaction of hevein with $(GlcNAc)_4$, indicating that a minimum sugar length ($n = 5$) is needed to allow the binding of two hevein domains. Therefore, we generated model structures of the possible 2:1 complexes, assuming that no significant distortion of the polypeptide chain takes place with respect to its conformation in 1:1 complexes (Figure S8). There are only three different possibilities for $(GlcNAc)_5$: the GlcNAc residues located at position +1 of both hevein molecules are at relative positions 1:3, 1:4 and 1:5 on the sugar chain. The recognition of two consecutive GlcNAc units at +1 of both domains would lead to severe steric interactions between the two polypeptides. Thus, the structures in Figure S8 represent the most possible mode of binding of hevein domains to a $(GlcNAc)_n$ chain and provide evidence for a multivalent interaction in nature between hevein and chitin.

Binding affinity

With the structure and the thermodynamic balance of the interaction data at hand, we can envisage two different origins to account for the observed increase in macroscopic affinity of the protein for $(GlcNAc)_5$ compared with $(GlcNAc)_4$. First, the pentasaccharide provides a larger number of contacts to the protein and therefore an increase in affinity would be expected. Second, to some extent, two protein molecules can be bound to $(GlcNAc)_5$. The ultracentrifugation data show, however, that the 2:1 stoichiometry is not favored in the presence of a slight excess of ligand, suggesting a rather low stability for the 2:1 complexes in comparison with the 1:1 ones. In addition, the models obtained for the 2:1 complexes indicate that for this short chitin fragment none of the individual proteins would retain the high number of protein–sugar contacts observed for the 1:1 complexes. It seems that each of the individual hevein–$(GlcNAc)_5$ interactions in the ternary complexes are therefore weaker than the corresponding ones in the 1:1 complexes, in which $(GlcNAc)_5$ extends over hevein surface.

The thermodynamic data indicate that the entropic term has a principal role in the affinity increase ($\Delta(T\Delta S) = T\Delta S(GlcNAc)_5 - T\Delta S(GlcNAc)_4 = 2.1$ kcal/mol). A similar behavior has been reported for WGA [28] and UDA [45] (Figure S9). In these cases, more favorable entropic balances were observed for $(GlcNAc)_5$ than for $(GlcNAc)_3$ binding, with $\Delta(T\Delta S)$ values of 1.5 kcal/mol (WGA), and 2.0 kcal/mol (UDA). For both UDA and WGA, however, the more favorable entropic term for $(GlcNAc)_5$ was partially counterbalanced by a less favorable enthalpic term, leading to a very small change in ΔG, in contrast with the observed behavior for hevein, for which $(GlcNAc)_5$ binding is also more enthalpically favorable than that of $(GlcNAc)_{3–4}$ (Table 1). Further studies will be necessary to interpret these data. Although merely speculative, this more favorable entropic balance for longer chitin fragments might reflect the existence of a number of possible binding modes. In addition, both sugar/protein flexibility and/or solvation affects could have a contribution to the observed entropic term [46].

Significance

The interaction of hevein with $(GlcNAc)_n$ oligomers, with n = 2–8 has been studied using a pluridisciplinary approach with isothermal titration calorimetry (ITC), ultracentrifugation and NMR to get a deeper understanding of the structural basis for chitin recognition by plant defense proteins. The first detailed experimental NMR structure of a protein–chitin complex is presented. In contrast to previous beliefs, our data show that the binding of GlcNAc units away from the nonreducing end is indeed possible. Moreover, the extension of the saccharide chain to positions –1 and –2 generates additional protein–carbohydrate contacts, further contributing to the complex stability. Hevein domains show an extended binding site which comprises not only residues 21, 23, 30 and 19, but also the loop region 13–16. Five to seven GlcNAc units in the ligand are required to express all the possible protein–sugar interactions simultaneously. The extended binding site of hevein domains is perfectly pre-organized for the exclusive recognition of linear β(1→4)-linked GlcNAc polymers when they adopt a conformation close to the global minimum. The conformation of the chitin fragment in the binding site is therefore essentially identical to the most populated one in the free state. A relayed in–out mechanism takes place so that the different GlcNAc units of the polysaccharide exchange their positions in the binding site by interacting with different protein subsites in a dynamic process. This event, together with the multiple binding of several hevein molecules to the same chitin chain, is probably at the origin of the biological activity of hevein. Chitin recognition by a large number of hevein molecules would be expected to have a net effect on the dynamical behavior of the polymer, modifying its physical properties. Through this mechanism, plant defense proteins could block the normal development of pathogen organisms, as fungus and insects, by interfering with the chitin biological function of protection.

Materials and methods

Source of lectin and ligand

Hevein was isolated from *Hevea brasiliensis* latex as described previously [30]. Hevein concentration was measured by UV absorbance ($\varepsilon = 12{,}100$). $(GlcNAc)_n$ oligomers were purchased from Toronto Research Chemicals Co. ^{13}C-labeled N,N′,N″,N‴,N⁗-pentacetyl-chitopentaose was a gift from E. Samain and H. Driguez in CERMAV-CNRS, Grenoble [47]

Microcalorimetry experiments

The calorimetric titrations were performed with a MES Microcal titration calorimeter, as described [48]. In particular, microliter amounts of the ligand solutions were added by means of a rotating stirrer-syringe to the protein solution contained in a 1.35 ml cell. Additional titrations were carried out in the pentasaccharide case. In these experiments, microliter amounts of the protein solutions were added to the ligand solution contained in a 1.35 ml cell. The total heat effect for each injection was corrected for the heat of dilution. The thermodynamic parameters of binding were obtained by analyzing the data with the software provided by Microcal Inc. The cumulative heat effect (Q) during the titration process for a simple set of binding sites is given by:

$$Q = M_t V_o n v \Delta H$$

Where M_t is the macromolecule concentration in the calorimetric cell characterized by the working volume V_o, n is the number of binding sites in the given set with a binding enthalpy of ΔH, and v is the fractional saturation of each type of site which can be related to the apparent association constant (K') and to the total ligand concentration (L_T):

$$K' = v / [(1-v) L]$$

$$L_T = L + M_t n v$$

Analytical ultracentrifugation experiments – sedimentation equilibrium

The experiments were performed using a Beckman Optima XL-A analytical ultracentrifuge equipped with absorbance optics, using an An50Ti rotor. As a first step, hevein was equilibrated using buffer 100 mM sodium chloride and 10 mM sodium phosphate, pH 5.6. The influence of oligosaccharide binding was studied by adding the sugars to a final concentration of 3.4–122.4 μM. Short column experiments (with 70 μl of hevein and loading concentrations between 17–69 μM, with and without sugar),were done using two consecutive speeds (28,000 and 36,000 rpm), by taking absorbance scans (0.001 cm step size and 10 averages) at the appropriate wavelength (280 or 290 nm) and at sedimentation equilibrium. We used six-channel centerpieces of charcoal-filled Epon (12 mm optical path). The equilibrium temperature was 25°C for most of the experiments. In addition, the experiments for the hevein–pentasaccharide complex were repeated also at 5°C and with a protein concentration of 1 mM in order to reproduce the conditions of the NMR experiments. In this case, double-sector centerpieces of charcoal-filled Eppon (3 mm optical path) were used. In all the cases, high-speed sedimentation (60,000 rpm) was afterwards conducted for baseline correction.

Whole-cell buoyant molecular masses ($M^c_{w,a}$) were determined by fitting a sedimentation equilibrium model for a single sedimenting solute to individual datasets with the program EQASSOC (supplied by Beckman; [49]). The partial specific volume of hevein was 0.68 ml/g, calculated from its amino-acid composition. The monomer relative molecular mass was taken as 4727.

NMR spectroscopy

NMR experiments. General aspects

^1H-NMR spectra of the hevein–pentasaccharide complex were recorded in 85:15 ^1H_2O:^2H_2O on Bruker DRX-500 and Varian Unity 500 MHz spectrometers at 5°C. Spectra for the 2D experiments were acquired using 1 mM solutions of protein. The TOCSY [50] and NOESY [51] experiments of the hevein–pentasaccharide complex were performed in the phase-sensitive mode using the TPPI method [52] for quadrature detection in F_1. Typically, a data matrix of $512 \times 2K$ points was used to digitize a spectral width of 7800 Hz. Eighty scans were used per increment with a relaxation delay of 1 s. Before Fourier transformation, zero filling was used in F_1 to expand the data to $1K \times 2K$. Baseline correction was applied in both dimensions. The corresponding shift was optimized for the different spectra. The TOCSY spectra were recorded using MLEV-17 [50] during the 60 ms of isotropic mixing period. The NOESY experiments were performed with mixing times of 50, 150 and 200 ms.

HSQC, HSQC–TOCSY and HSQC–NOESY experiments were carried out to obtain the complete ^1H and ^{13}C assignment of the ^{13}C labeled pentasaccharide. For the free state, the one-bond proton–carbon correlation experiment was collected using the gradient-enhanced HSQC sequence. A data matrix of $1K \times 1K$ was used to digitize a spectral width of 3000 Hz in F_2 and 15,000 Hz in F_1. 4 scans were used per increment with a relaxation delay of 1 s and a delay corresponding to a J value of 145 Hz. ^{13}C decoupling was achieved by

the WALTZ scheme. The 2D–HSQC–TOCSY experiment was conducted with 80 ms of mixing time (MLEV 17). The same conditions as for the HSQC were used with eight scans. The 2D HSQC–NOESY experiment was conducted with a mixing time of 500 ms and 32 scans.

NMR titration experiments

The binding of *N*-acetyllactosamine to hevein was monitored by recording 1D 500 MHz ^{1}H-NMR spectra of a series of samples with variable sugar concentration (ten different concentrations) at 25°C. 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 1.0 ml in buffer (85:15, ^1H_2O:^2H_2O, 100 mM sodium chloride and 10 mM sodium phosphate, pH 5.6). The concentration of hevein was calculated from its UV absorbance at 280 nm. The 1D NMR spectrum for the sample with the highest ligand:protein ratio was recorded by dissolving the sugar (~47 mM) in 0.5 ml of the hevein solution described above. The sample with the other 0.5 ml of this hevein solution was used to obtain the ^{1}H-NMR chemical shifts of the free-sugar protein sample (δ_{free}). The titration curve was obtained by adding small aliquots of the highest ligand:protein ratio sample to the ligand-free protein one as described [23,24].

NMR analysis of the hevein- N,N′,N″,N‴,N⁗-pentacetyl-chitopentaose complex. General considerations.

Owing to the existence of at least two complexes in solution, the NMR analysis was not trivial and will be detailed for sake of clarity. As a first step, the complete ^{1}H and ^{13}C assignment of the ^{13}C-labeled pentasaccharide in the free state was carried out by using HSQC, HSQC-TOCSY and HSQC-NOESY experiments. The NMR spectra of hevein and its complexes with $(GlcNAc)_n$, with $n = 1–3$, have been previously analyzed by us [23,24]. Thus, these spectra were taken as starting points for the evaluation of the more complicated ones obtained for pentasaccharide binding. In fact, different protein:ligand ratios had to be used to get the maximum information. In particular, we used six different samples with protein:ligand molar ratios ranging from 2:1 to 1:4. Most of the structural information was derived from the analysis of NOESY data obtained for a 1:4 molar ratio. Under these conditions, the 1:1 stoichiometry is very predominant in solution (95%), according to analytical ultracentrifugation experiments carried out under the exact NMR conditions (5°C and 1 mM protein solutions). NOESY and ROESY spectra showed the existence of at least two 1:1 protein–sugar complexes in solution corresponding to the specific recognition of different GlcNAc units of the pentasaccharide at subsite +1 ([24], Figure S2). In general, the use of relatively short $(GlcNAc)_n$ fragments as chitin models has the advantage of providing low molecular weight complexes that can be analyzed by NMR methods. In fact, it would have been impossible to collect reasonable NMR data for the protein–sugar complexes by using the natural target, chitin. However, as the recognition of other GlcNAc units apart from the nonreducing end at subsite +1 is indeed possible (see the Results and discussion, as deduced in the present work from NMR, ITC and analytical ultracentrifugation experiments), short $(GlcNAc)_n$ oligomers can produce a mixture of non-equivalent protein–sugar complexes in solution. With respect to the pentasaccharide, the fact that all the possible hevein–$(GlcNAc)_5$ 1:1 complexes are not totally equivalent is not relevant from the biological point of view. Indeed, for a long $(GlcNAc)_n$ chain or for chitin itself, a mixture of totally equivalent hevein–sugar complexes, both from the structural and thermodynamic point of view, should be expected to occur. Therefore, all the structural information obtained from the NMR analysis of the hevein–$(GlcNAc)_5$ complexes was used to generate a 3D model of the biologically relevant hevein–chitin complex.

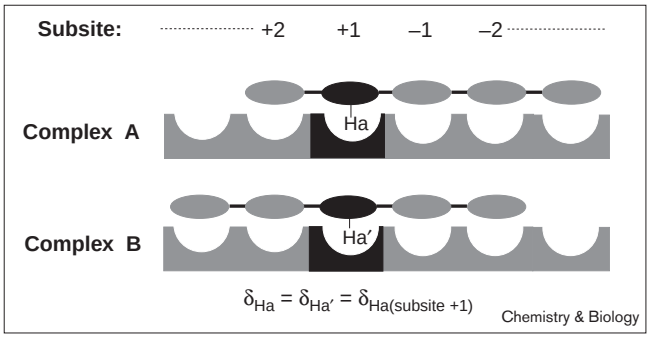
Assignment

The fact that the NMR data corresponds to a mixture of at least two 1:1 complexes in solution, along with the extreme overlapping of the pentasaccharide signals in the free state, posed a great challenge on the assignment of the sugar signals in the bound state. Under the experimental conditions used (protein:ligand, 1:4 molar ratio), free–bound exchange peaks were observed for many sugar signals. In a first stage, the exchange signals cannot be directly used for assignment, because of the extreme overlapping observed in the spectrum corresponding to the free pentasaccharide (the three middle GlcNAc units are degenerated). Therefore we used an iterative protocol. First, sugar signals in the bound state were classified according to the proton type (i.e. H1, H2..., and so on) by using those free–bound exchange sugar peaks, observed in ROESY/NOESY spectra at 5°C, in combination with HSQC-type experiments. These combinations were crucial, as the ^{13}C-NMR chemical shift are well dispersed.

In some cases, clear exchange peaks between different signals in the bound state corresponding to the same proton type (i.e. H1 protons with three different chemical shifts, H2 protons with more than two different chemical shifts..., and so on) can also be observed in the ROESY/NOESY spectra. For example, four different signals (δ_{boundA}, δ_{boundB}, δ_{boundC}, δ_{boundD}) corresponding to the bound states can be observed for the sugar methyl acetamide moieties. ROESY spectra clearly show that the signal with chemical shift δ_{boundB} is in exchange with the signal at δ_{boundD} (ppm) and with the free species (δ_{free}; the five methyl chemical shifts are degenerated). Thus, δ_{boundB} may represent more than one methyl group of the pentasaccharide, but this particular chemical shift labels these methyl groups when they are influenced by the structural environment of subsite B within the complex. Therefore, every signal corresponding to the bound ligand represents a certain disposition of a particular proton type within the complexes, and it is only necessary to label the signals according to the proton type (i.e. H1, H2..., and so on), and to a given subsite (H1 at subsite +1, H1 at subsite −1, H1 at subsite −1 ...; see Figure 11). The subsite assignment was carried out in the first place by using the observed protein–sugar NOEs. Only the unambiguous protein–sugar NOEs were used, along with the observed protein–protein NOEs, to generate ten initial 3D models of the chitin–hevein complex. Our NMR structure of the protein in the complex with methyl-β-chitobioside [53] was used as starting structure in a simulated annealing protocol [53] with the AMBER force field [54], as implemented into the DISCOVER 2.9 program (Biosym Technologies, USA). The experimentally observed protein–carbohydrate NOEs were added as upper-bound constraints (5 Å). Given the limited flexibility around the glycosidic linkages of the pentasaccharide (typical of β(1→4) linear oligosaccharides [53]), these loose constraints were sufficient to define the position of the ligand in the binding site.

After an initial restrained energy minimization (REM) with 2000 conjugate gradient iterations, the structures were equilibrated at 600 K for 2 ps, and, at this temperature, their conformational behavior for the next

Figure 11



Representation of different 1:1 hevein:$(GlcNAc)_5$ complexes. The chemical shift of proton Ha of the second (δa), third (δa′), and fourth (δa″) GlcNAc units in subsite +1 is identical. This chemical shift can therefore be considered as representative of any proton Ha in subsite +1 and its assignment to any particular GlcNAc unit is not required.

2 ps was simulated by restrained molecular dynamics (RMD). In the next step, the structures were then subjected to a cooling regime, in which the temperature was decreased in 100 K every 2 ps until a temperature by 100 K was reached. At this temperature, 4 ps of RMD calculation were carried out. The final structures underwent REM using 2000 conjugate gradient iterations. In a second step, and once the structures were at hand, the theoretical $\Delta\delta$ values induced in the sugar protons as a result of complexation were calculated from these initial models and compared with those experimentally derived (exchange peaks from the NOESY/ROESY spectra). The Johnson–Bovey model [55] was used, as implemented in the program MOLMOL. Both ring current and bond polarization effects were taken into account. The theoretical $\Delta\delta$ values were derived from the ten protein–sugar complexes and then averaged. This calculation confirmed all the sugar assignments carried out in the first stage. In addition, it allowed the assignment of several new sugar signals which in turn permitted the detection of additional protein–sugar NOEs. These additional constraints were further used to refine the complex. A final model of the biologically relevant hevein–chitin complex was generated, using 350 protein–protein NOEs and 22 unambiguous protein–sugar NOEs. The sugar $\Delta\delta$ values derived from these final set of structures were in good agreement with those obtained experimentally.

With regard to the ligand, its conformation in the free state was analyzed by standard NMR methods and molecular mechanics calculations. As this is a linear $\beta(1\rightarrow4)$-linked trisaccharide, with no possible contacts between non consecutive units, the calculated (from the relaxed energy surfaces) global minimum of the disaccharide *N,N′*-diacetylchitobiose [44] was used to generate the starting geometry for every glycosidic linkage, as previously described [23,24]. Molecular dynamics simulations were then carried out to access the conformational fluctuations around the linkages. The obtained geometry was used to generate a longer chitin model and submitted to a molecular dynamics simulation protocol (200 ps) at 300 K, using the AMBER force field parametrized for carbohydrates [56]. After minimization of the resulting structures, a representative geometry was chosen to model the bound polymer. This chosen conformation was found to be in perfect agreement with the observed protein–sugar contacts.

Supplementary material

Supplementary material including Table S1–S5 and Figures S1–S9 with details about the ITC, analytical ultracentrifugation and NMR analysis of $(GlcNAc)_{2-8}$ binding by hevein is available at http://current-biology.com/supmat/supmatin.htm.

Acknowledgements

Financial support from DGICYT (grant PB-96-0833) and from the Acciones Integradas program is gratefully acknowledged. We thank Martín-Lomas for his interest and support throughout this work. J.L.A. thanks CAM and MEC for fellowships. We thank E. Samain and H. Driguez from CERMAV-CNRS (Grenoble) for a gift of the ^{13}C-labeled pentasaccharide. We also thank C.A. Botello and G. Rivas (CIB-CSIC) for the equilibrium sedimentation experiments, and M. Menéndez (IQFR-CSIC) for discussions regarding the ITC data, and J.J. Beintema (Gromingen) for hevein supply.

References

1. Varki, A. (1993). Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97-130.
2. Lasky, L.A. (1992). Selectins: interpreters of cell-specific carbohydrate information during inflammation. *Science* **258**, 964-9.
3. Dwek, R.A. (1996). Glycobiology. *Chem. Rev.* **96**, 683-720.
4. Montreuil, J., Vliegenthart, J. & Schachter, H. (eds) (1995). *Glycoproteins*. Elsevier, Amsterdam.
5. Gabius, H.J. (1997). Animal lectins. *Eur. J. Biochem.* **243**, 543-576.
6. Gabius, H.J. (1997). Concepts of tumor lectinology. *Cancer Invest.* **15**, 454-464.
7. Kaltner, H. & Stierstorfer, B. (1998). Animal lectins as cell adhesion molecules. *Acta Anat.* **161**, 162-179.
8. Reuter, G. & Gabius, H.J. (1999). Eukaryotic glycosylation: whim of nature or multipurpose tool? *Cell. Mol. Life Sci.* **55**, 368-422.
9. Rudiger, H. (1998). Plant lectins – more than just tools for glycoscientists: occurrence, structure, and possible functions of plant lectins. *Acta Anat.* **161**, 130-152.
10. Rudiger, H., *et al.,* & Gabius, H.J. (2000). Medicinal chemistry based on the sugar code: fundamentals of lectinology and experimental strategies with lectins as targets. *Curr. Med. Chem.* **7**, 389-416.
11. Vyas, N.K. (1991). Atomic features of protein-carbohydrate interactions. *Curr. Opin. Struct. Biol.* **1**, 732-740.
12. Weis, W.I. & Drickamer, K. (1996). Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **65**, 441-473.
13. Loris, R., Hamelryck, T., Bouckaert, J. & Wyns, L. (1998). Legume lectin structure. *Biochim. Biophys. Acta* **1383**, 9-36.
14. Bush, C.A., Martin-Pastor, M. & Imberty, A. (1999). Structure and conformation of complex carbohydrates of glycoproteins, glycolipids, and bacterial polysaccharides. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 269-293.
15. Quiocho, F.A. (1986). Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. *Annu. Rev. Biochem.* **55**, 287-315.
16. Imberty, A., Bourne, Y., Cambillau, C., Rouge, P. & Perez, S. (1993). Oligosaccharide conformation in protein/carbohydrate complexes. *Adv. Biophys. Chem.* **3**, 71-118.
17. Peters, T. & Pinto, B.M. (1996). Structure and dynamics of oligosaccharides: NMR and modeling studies. *Curr. Opin. Struct. Biol.* **6**, 710-720.
18. von der Lieth, C., *et al.*, & Gabius, H. (1998). Lectin ligands: new insights into their conformations and their dynamic behavior and the discovery of conformer selection by lectins. *Acta Anat.* **161**, 91-109.
19. Richardson, J.M., Evans, P.D., Homans, S.W. & Donohue-Rolfe, A. (1997). Solution structure of the carbohydrate-binding B-subunit homopentamer of verotoxin VT-1 from *E. coli*. *Nat. Struct. Biol.* **4**, 190-193.
20. Poveda, A. & Jimenez-Barbero, J. (1998). NMR studies of carbohydrate–protein interactions in solution. *Chem. Soc. Rev.***27**, 133-144.
21. Jimenez-Barbero, J., Asensio, J.L., Canada, F.J. & Poveda, A. (1999). Free and protein-bound carbohydrate structures. *Curr. Opin. Struct. Biol.* **9**, 549-555.
22. Raikhel, N.V., Lee, H.I. & Broekaert, W.F. (1993). Structure and function of chitin-binding proteins. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **44**, 591-615.
23. Asensio, J.L., Canada, F.J., Bruix, M., Rodriguez-Romero, A. & Jimenez-Barbero, J. (1995). The interaction of hevein with *N*-acetylglucosamine-containing oligosaccharides. Solution structure of hevein complexed to chitobiose. *Eur. J. Biochem.* **230**, 621-633.
24. Asensio, J.L., *et al.*, & Jimenez-Barbero, J. (1998). NMR investigations of protein-carbohydrate interactions: refined three- dimensional structure of the complex between hevein and methyl beta-chitobioside. *Glycobiology*, **8**, 569-577.
25. Martins, J., Maes, D., Loris, R., Pepermans, H.A.M., Wyns, L., Willen, R. & Verheyden, P. (1996). NMR study of the solution structure of Ac-AMP2, a sugar binding antimicrobial protein isolated from *Amaranthus caudatus*. *J. Mol. Biol.* **258**, 322-333.
26. Kronis, K.A. & Carver, J.P. (1985). Thermodynamics of wheat germ agglutinin-sialyloligosaccharide interactions by proton nuclear magnetic resonance. *Biochemistry* **24**, 834-840.
27. Kronis, K.A. & Carver, J.P. (1985). Wheat germ agglutinin dimers bind sialyloligosaccharides at four sites in solution: proton nuclear magnetic resonance temperature studies at 360 MHz. *Biochemistry* **24**, 826-833.
28. Bains, G., Lee, R.T., Lee, Y.C. & Freire, E. (1992). Microcalorimetric study of wheat germ agglutinin binding to N- acetylglucosamine and its oligomers. *Biochemistry* **31**, 12624-12628.
29. Garcia-Hernandez, E., Zubillaga, R.A., Rojo-Dominguez, A., Rodriguez-Romero, A. & Hernandez-Arana, A. (1997). New insights into the molecular basis of lectin-carbohydrate interactions: a calorimetric and structural study of the association of hevein to oligomers of *N*-acetylglucosamine. *Proteins Struct. Funct. Genet.* **29**, 467-477.
30. Siebert, H.C., *et al.*, & Gabius, H.J. (1997). Role of aromatic amino acids in carbohydrate binding of plant lectins: laser photo chemically induced dynamic nuclear polarization study of hevein domain-containing lectins. *Proteins Struct. Funct. Genet.* **28**, 268-284.
31. Wright, C.S. (1984). Structural comparison of the two distinct sugar binding sites in wheat germ agglutinin isolectin II. *J. Mol. Biol.* **178**, 91-104.
32. Wright, C.S. (1990). 2.2 A resolution structure analysis of two refined *N*-acetylneuraminyl- lactose–wheat germ agglutinin isolectin complexes. *J. Mol. Biol.* **215**, 635-651.

33. Wright, C.S. (1992). Crystal structure of a wheat germ agglutinin/ glycophorin-sialoglycopeptide receptor complex. Structural basis for cooperative lectin-cell binding. *J. Biol. Chem.* **267**, 14345-14352.
34. Lee, R.T. & Lee, Y.C. (1997). *Neoglycoconjugates*. Chapman and Hall, London.
35. Gordon, E.J., Gestwicki, J.E., Strong, L.E. & Kiessling, L.L. (2000). Synthesis of end-labeled multivalent ligands for exploring cell-surface-receptor–ligand interactions. *Chem. Biol.* **7**, 9-16.
36. Gordon, E.J., Sanders, W.J. & Kiessling, L.L. (1998). Synthetic ligands point to cell surface strategies. *Nature* **392**, 30-31.
37. Mammen, M., Seok-Ki, C. & Whitesides, G.M. (1998). Polyvalent interactions in biological systems: implications for design and use of multivalent ligands and Inhibitors. *Angew. Chem. Int. Ed.* **37**, 2754-2794.
38. Allen, A.K., Neuberger, A. & Sharon, N. (1973). The purification, composition and specificity of wheat-germ agglutinin. *Biochem. J.* **131**, 155-162.
39. Carver, J.P., Michnick, S.W., Imberty, A. & Cumming, D.A. (1991). Oligosaccharide–protein interactions: a three dimensional view. *Ciba Found. Symp.* **158**, 6-26.
40. Searle, M.S. & Williams, D.H. (1992). The cost of conformational order: entropy changes in molecular associations. *J. Am. Chem. Soc.* **114**, 10690-10697.
41. Gabius, H.J. (1998). The how and why of protein-carbohydrate interaction: a primer to the theoretical concept and a guide to application in drug design. *Pharm. Res.* **15**, 23-30.
42. Jeffrey, G.A. & Saenger, W. (1991). *Hydrogen Bonding in Biological Structures*. Springer-Verlag, Berlin.
43. Paz, M.L.d.l., Jiménez-Barbero, J. & Vicent, C. (1998). Hydrogen bonding and cooperativity effects on the assembly of carbohydrates. *Chem. Commun.* 465.
44. Espinosa, J.F., Asensio, J.L., Bruix, M. & Jimenez-Barbero, J. (1996). Solution conformation of chitobiose. Evidences for the existence of conformational averaging from NMR spectroscopy measurements and molecular mechanics calculations. *An. Quim.* **92**, 320-324.
45. Lee, R.T., Gabius, H.J. & Lee, Y.C. (1998). Thermodynamic parameters of the interaction of Urtica dioica agglutinin with *N*-acetylglucosamine and its oligomers. *Glycoconj. J.* **15**, 649-655.
46. Zidek, L., Novotny, M.V. & Stone, M.J. (1999). Increased protein backbone conformational entropy upon hydrophobic ligand binding. *Nat. Struct. Biol.* **6**, 1118-1121.
47. Samain, E., Drouillard, S., Heyraud, A., Driguez, H. & Geremia, R.A. (1997). Gram-scale synthesis of recombinant chitooligosaccharides in *Escherichia coli*. *Carbohydr. Res.* **302**, 35-42.
48. Wiseman, T., Williston, S., Brandts, J.F. & Lin, L.N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179**, 131-137.
49. Minton, A.P. (1994). In *Modern Analytical Ultracentrifugation*. Birkhauser, Boston. Conservation of sigmal: A new algorithm for the elimination of the reference concentration as an independently variable parameter in the analysis of sedimentation equilibrium. pp.81-93.
50. Bax, A. & Davis, D.G. (1985). MLEV-17 based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **65**, 355-360.
51. Kumar, A., Ernst, R.R. & Wuthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95**, 1-6.
52. Marion, D. & Wuthrich, K. (1983). Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of 1H-1H spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Commun.* **113**, 967-74.
53. Scheek, R.M., van Gunsteren, W.F. & Kaptein, R. (1989). Molecular dynamics simulation techniques for determination of molecular structures from nuclear magnetic resonance data. *Methods Enzymol.* **177**, 204-218.
54. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. & Weiner, P.J. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, **106**, 765-776.
55. Case, D.A. (1995). Calibration of ring-current effects in proteins and nucleic acids. *J. Biomol. NMR* **6**, 341-346.
56. Homans, S.W. (1990). A molecular mechanical force field for the conformational analysis of oligosaccharides: comparison of theoretical and crystal structures of Man α 1-3Man β 1-4GlcNAc. *Biochemistry* **29**, 9110-9118.