Biochemical and Structural Analysis of *Helix pomatia* Agglutinin

**A HEXAMERIC LECTIN WITH A NOVEL FOLD***

Received for publication, April 11, 2006, and in revised form, May 15, 2006. Published, JBC Papers in Press, May 16, 2006. DOI 10.1074/jbc.M603452200

Jean-Frederic Sanchez 1,2, Julien Lescar 3,5, Valérie Chazalot, 2 Ayméric Audfray, 2 Jean Gagnon, 6, Richard Alvarez, 1 Christelle Breton, 3 Anne Imbert, 1,6 and Edward P. Mitchell 1,2, 4

From the 1Centre de Recherches sur les Macromolécules Végétales, CNRS (affiliated with Université Joseph Fourier), 38041 Grenoble, France, 5School of Biological Sciences, Nanyang Technological University, 637616 Singapore, 6CNRS-FRE2685, Université Joseph Fourier, 38706 La Tronche, France, 7Department of Biochemical/Molecular Biology, University of Oklahoma, Oklahoma City, Oklahoma 73104, and **European Synchrotron Radiation Facility, Experiments Division, 38043 Grenoble, France.

*Helix pomatia* agglutinin (HPA) is a N-acetylgalactosamine (GalNAc) binding lectin found in the albumen gland of the roman snail. As a constituent of perivitelline fluid, HPA protects fertilized eggs from bacteria and is part of the innate immunity system of the snail. The peptide sequence deduced from gene cloning demonstrates that HPA belongs to a family of carbohydrate-binding proteins recently identified in several invertebrates. This domain is also present in discoidin from the slime mold *Dictyostelium discoideum*. Investigation of the lectin specificity was performed with the use of glycan arrays, demonstrating that several GalNAc-containing oligosaccharides are bound and rationalizing the use of this lectin as a cancer marker. Titration microcalorimetry performed on the interaction between HPA and GalNAc indicates an affinity in the 10^-4 M range with an enthalpy-driven binding mechanism. The crystal structure of HPA demonstrates the occurrence of a new B-sandwich lectin fold. The hexameric quaternary state was never observed previously for a lectin. The high resolution structure complex of HPA with GalNAc characterizes a new carbohydrate binding site and rationalizes the observed preference for αGalNAc-containing oligosaccharides.

Because of their ability to recognize complex carbohydrates on cell surfaces with high specificity, lectins are proteins that play important roles in the social life of cells. A growing repertoire of lectins has been identified in invertebrates (1), where these molecules are involved in self/non-self recognition (2). Examples include the recognition of “sister” cells as part of the aggregation mechanisms in primitive organisms (e.g. slime molds (3), sponges and corals (4)), the specific binding of polysaccharide-coated pathogenic bacteria in the innate immunity system of invertebrates (5, 6), and the mediation of symbiosis, for example between coral and their symbiotic algae (7).

In the edible roman snail, a lectin named *Helix pomatia* agglutinin (HPA) 5 is secreted by the albumen gland that produces the perivitelline fluid, composed of protein and polysaccharide complexes that coat each fertilized egg (8). This lectin has the property to aggregate bacteria such as group C streptococci (9) and *Listeria monocytogenes* (10) and is also able to bind to the surface of the herpesvirus (11).

Despite the potential antimicrobial activity of HPA, biochemical and structural data are scarce. An agglutination activity was identified in an extract of snail albumen gland at the end of 19th century by its ability to aggregate red cells and human milk particles (12). Later on, the purified lectin was shown to agglutinate blood group A red cells (13), but not those of blood groups B or O (14). The binding preference was established to be Forssman antigen (αGalNAc1–3GalNAc-R) >blood group A substance (αGalNAc1–3[αFuc1–2]Gal) >Tn antigen (αGalNAc-Ser/Thr) >GalNAc >GlcnAc, therefore confirming the specificity for terminal α-N-acetyl-D-galactosamine (αGalNAc) residues (15). HPA was identified as a hexamer with a total molecular mass of 79 kDa (14). Further characterization led to the conclusion that the snail hemagglutinin consists of six identical polypeptide subunits, each containing one intra-chain disulfide bond and a carbohydrate binding site. In fact, two chains associate by a disulfide bridge, and the hexamer therefore consists of a non-covalent trimer of covalent dimers (16). HPA is a glycoprotein with ~7% carbohydrate content, and because this glycoprotein could be fractionated into at least 12 components by isoelectric focusing (17), the presence of different glycoforms is probable.

---

5 The abbreviations used are: HPA, *Helix pomatia* agglutinin; GalNAc, N-acetylgalactosamine; GlcNAc, N-acetylgalcosamine; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; RACE, rapid amplification of cDNA ends; GD2, GalNAcβ1→4[NeuAcα2→8NeuAcα2→3]Galβ1→4Glcβ1-Cer.
**Crystal Structure of Helix pomatia Lectin**

**TABLE 1**

<table>
<thead>
<tr>
<th>Data collection and phasing statistics</th>
<th>HPA native</th>
<th>HPA-GalNAc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution limits (Å)</td>
<td>53.4-2.40</td>
<td>11.2-1.30</td>
</tr>
<tr>
<td>Working set R (%) (observations)</td>
<td>22.2 (4785)</td>
<td>14.43 (30995)</td>
</tr>
<tr>
<td>Test set R (%) (observations)</td>
<td>26.6 (236)</td>
<td>17.90 (1643)</td>
</tr>
<tr>
<td>Average Rmerge (%)</td>
<td>8.4 (54.2)</td>
<td>6.9 (38.0)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.6 (100.0)</td>
<td>99.8 (99.9)</td>
</tr>
<tr>
<td>Wilson B-factor (Å²)</td>
<td>53.0</td>
<td>56.3</td>
</tr>
<tr>
<td>Phasing power</td>
<td>0.797 (2.389)</td>
<td>n/a</td>
</tr>
<tr>
<td>Average figure of merit</td>
<td>0.884 (0.487)</td>
<td>n/a</td>
</tr>
<tr>
<td>Number of residues/protein atoms</td>
<td>100/784</td>
<td>99/819</td>
</tr>
<tr>
<td>Number of solvent/sugar/zinc/other ligands</td>
<td>24/0/0/0</td>
<td>99/29/1/14</td>
</tr>
<tr>
<td>Number of residues in alternative conformations</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Number of outliers on Ramachandran plot</td>
<td>92% most favored, 94% most favored, 1 disallowed (Ser-39)</td>
<td>0 disallowed (Ser-39)</td>
</tr>
<tr>
<td>Overall G-factor from Procheck</td>
<td>1.8</td>
<td>–0.4</td>
</tr>
<tr>
<td>R.m.s.d. distance and angle deviations</td>
<td>0.012</td>
<td>0.013</td>
</tr>
<tr>
<td>Bond distances (Å)</td>
<td>1.46</td>
<td>0.32</td>
</tr>
<tr>
<td>Planar angles (°)</td>
<td>0.004</td>
<td>0.032</td>
</tr>
</tbody>
</table>

**EXPERIMENTAL PROCEDURES**

**Materials**—HPA was purchased from Sigma and Vector (Burlingame, CA). Oligonucleotides for PCR were purchased from MWG SA, Biotech (Courtaboeuf, France). One H. pomatia snail was kindly donated by the snail farm, Escargot du Pont Chardon (Chevrieres, France).

**Peptide Sequencing and Mass Spectrometry**—CNBr fragments of HPA were prepared by addition of a 10-fold by weight excess of CNBr to the protein (10 mg/ml) in 70% HCOOH, and the mixture was kept in the dark for 20 h at 4 °C. The mixture was then diluted 1:10 with water and freeze-dried. The digest was dissolved in 6 M GdnHCl, pH 2.5, and separated by high performance liquid chromatography on a C4 column (2.1 × 100 mm) using a gradient of 0.1% trifluoroacetic acid and 70% CH3CN containing 0.085% trifluoroacetic acid. Amino acid sequence determination of collected fractions was performed using a Procise Sequencer model 492 (Applied Biosystems).

Phenythiohydantoin amino acid derivatives were identified and quantified by high performance liquid chromatography analysis on-line as recommended by the manufacturer. MALDI-TOF spectra were recorded on a Bruker Autoflex (Centre de Recherche sur les Macromolécules Végétales) and Applied Biosystems Voyager 125 (Institut de Biologie Structurale).

**Reverse Transcription PCR and cDNA Cloning of Helix pomatia Agglutinin**—PCR was used to amplify the coding region of HPA using degenerated primers that were designed with regard to the N-terminal sequence of the mature protein as deduced from the high resolution crystal structure of the HPA-GalNAc complex. Total RNA was obtained by grinding one albumen gland in liquid nitrogen followed by an extraction using TRI reagent (Sigma). cDNAs were synthesized using rapid amplification of cDNA ends (3′-RACE kit; Invitrogen) with the following primers: pr-hpa (sense 5′-TGGTGGAAGAYGAYCNGGNTGG-3′), which codes for the peptide sequence CGNDAGW, and UAP (Universal Amplification Primer; Invitrogen). PCR was performed using cDNA as the template and Taq DNA polymerase (Promega) with the following temperature profile: 4 min at 95 °C first and then 45 s at 95 °C, 1 min at 45 °C, 5 min at 72 °C, for 39 cycles. These conditions permitted the amplification of a fragment of 650 bp. The purified PCR fragment was cloned into the pGEM-T easy vector (Promega). Five independent clones were sequenced on both strands (Genome Express, Meylan, France). The 5′-end of the sequence was obtained using the 5′-RACE kit from Invitrogen. cDNA was synthesized using an HPA gene-specific antisense oligonucleotide (GSP1-HPA: 5′-TCTATGACATGGCAAGACG-3′). PCR was then performed using this cDNA preparation as the template, Taq DNA polymerase, and the primers AAP (Abridged Anchor Primer; Invitrogen) and a second HPA gene-specific primer GSP2-HPA (antisense 5′-TGAGAATTCCCT-3′). PCR was performed with the following cycling temperature profile: 45 s at 94 °C, 1 min at 50 °C, 2 min at 72 °C, for 35 cycles. The purified PCR fragment
was ligated in the pGEM-T easy vector (Promega). Six independent clones were sequenced on both strands.

**Crystallization and Data Collections**—Crystallization trials were performed with Hampton crystallization screens I and II (Hampton Research, Laguna Niguel, CA). After optimization, crystals were obtained using the hanging drop method with precipitation solution containing lithium sulfate (2 M) and ammonium sulfate (3.5 M) in sodium citrate buffer (1.5 M, pH 6.5). Drops were made of an equal volume of reservoir solution and protein solution at 10 mg/ml. For the co-crystallization assays, the ligand was added to the protein solution with a 3-fold excess in molarity. Crystals were cryocooled at 100 K after soaking them for the shortest possible time in precipitant solution with 25% polyethylene glycol 3000 and 25% glycerol added. Data images were recorded on an ADSC Q4R CCD detector (Quantum

**FIGURE 1.** cDNA sequence and deduced amino acid sequence of HPA. The signal peptide sequence is indicated in *italics* and the sequence used for designing degenerated primer in *bold*. The three peptides sequenced after cyanogen bromide cleavage are *underlined* with different styles. The *N*-glycosylated asparagine is indicated by a *gray* background. Numbering of the protein starts with the first amino acid of mature protein.

**FIGURE 2.** Glycan array analysis of HPA as measured by fluorescence intensity. For clarity, part of the array is not shown (indicated by //) as no significant binding was observed for these sugars.
Crystal Structure of Helix pomatia Lectin

Corporation at beamline ID14–2 at the European Synchrotron Radiation Facility (ESRF) (Grenoble, France). Diffraction images were processed using MOSFLM (18) and scaled and truncated to structure factors using the CCP4 (19) programs SCALA and TRUNCATE. Data processing statistics are presented in Table 1. The presence of zinc in one batch of HPA was confirmed using a XANES spectrum on beamline ID14–4 at the ESRF.

Structure Determinations—The crystal structure of the HPA-GalNAc complex was determined using the single wavelength anomalous diffraction technique with highly redundant data from the crystals of the HPA-GalNAc-zinc complex to 1.7 Å resolution. Harker sections of the anomalous difference Patterson map showed peaks corresponding to one zinc atom/monomer in the asymmetric unit. The location of the zinc ion was found with HySS (20), and experimental phases were calculated and the position of the zinc refined with autoSHARP (21). A first model was built automatically with ARP/wARP (22) and used to prepare primers for cloning of the hpa gene and then completed using the obtained HPA sequence. Cycles of refinement with REFMAC (26) and manual rebuilding with O (24) resulted in an initial model of the lectin.

This model was used as the search probe for molecular replacement for the native protein and the GalNAc-zinc complex using MOLREP (25). In both the native structure (2.4 Å) and the high resolution HPA-GalNAc complex (1.3 Å), several amino acids missing in the search model could be added and sugar residues clearly defined in density were positioned manually using O (24). Cycles of refinement with REFMAC (26), including automatic water molecule placement using ARP/wARP (22), for the native structure and with ShelX (23) using anisotropic B-factors for the complex and manual rebuilding with O resulted in the final models. The refinement statistics are listed in Table 2. Stereochemical checks were performed with the PROCHECK program (27).

Glycan Microarray Analysis—HPA was labeled with Alexa Fluor 488-TPF (Invitrogen) according to the manufacturer’s instructions. The labeled protein was purified on a D-Salt polyacrylamide desalting column (Pierce). Alexa-labeled HPA was used to probe the printed glycan arrays (28) following the standard procedure of Core H of the Consortium for Functional Glycomics (www.functionalglycomics.org/).

Titration Microcalorimetry Measurements—ITC experiments were performed with a VP-ITC isothermal titration calorimeter (Microcal). The experiments were carried out at 25 °C. GalNAc (Sigma) and HPA were dissolved in the same buffer (0.1 M Tris-HCl, pH 7.5). Two experiments were conducted with HPA concentrations in the microcalorimeter cell (1.4478 ml) of 0.32 and 0.35 mM. GalNAc solution, in 30 injections of 10 μl at concentrations varying from 2.8 to 3.0 mM, was added at intervals of 5 min while stirring at 310 rpm. Control experiments performed by injections of buffer in the protein solution yielded insignificant heats of dilution. The experimental data were fitted to a theoretical titration curve using software supplied by Microcal with ΔH (enthalpy change), K_a (association constant), and n (number of binding sites/monomer) as adjustable parameters, from the classical relationship (29). All experiments were performed with c values 100 <c <200 (c = K_a × M where M is the initial concentration of the macromolecule).

RESULTS

Sequence Determination and Biochemical Characteristics—On SDS-polyacrylamide gel electrophoresis, under strong reducing conditions, the commercial HPA preparation showed a single band with an apparent molecular mass of 13 kDa, while the MALDI-TOF spectra (data not shown) indicated the presence of major peaks corresponding to a monomer (12.7 kDa), dimer (24.6 kDa), and hexamer (76 kDa) in agreement with previously published data (16). Cyanogen bromide cleavage and subsequent peptide sequencing led to the determination of three peptide sequences (Fig. 1). The sequence of a heptapeptide from the N-terminal region of the mature protein (CGND-AGW) was unambiguously deduced from the high resolution x-ray crystal structure of HPA-GalNAc complex, and this information allowed for cloning the corresponding hpa gene. Sequence analysis revealed that the gene encodes for a 121-amino acid polypeptide, including a peptide signal of 20 residues as deduced from prediction servers (CBS, www.cbs.dtu.dk/services/). The mature protein consists of 101 amino acids, with a calculated molecular mass of 11.3 kDa and an estimated pI of 8.20 (Fig. 1). The nucleotide sequence reported
here was deposited in GenBank™/EBI data bank with accession number DQ341310. The peptide sequence displays only 53% identity with an HPA sequence deposited by an independent study under accession number Q575S3 (discussed below).

HPA has been described as a glycoprotein, and one potential N-glycosylation site is present at Asn-34 (mature protein numbering). Mass spectrometry experiments (data not shown) indicated the existence of several glycoforms for the dimer, with three major peaks corresponding to molecular masses of 22,795, 23,967, and 25,139 Da. These peaks are interpreted as non-glycosylated, monoglycosylated, and diglycosylated dimers, respectively. Indeed, the difference of 1,172 Da observed between each form corresponds closely to the low molecular mass N-glycan (molecular mass 1,170 Da) that has been characterized as the

FIGURE 4. Ribbon representation of HPA crystal structure. A, monomer with the intramolecular disulfide bridge displayed as sticks. B, trimer represented along the 3-fold axis. C, hexamer with intermolecular disulfide bridges displayed as sticks. Plots are made using MOLSCRIPT (45) and RASTER-3D (46) unless otherwise indicated.
most represented on *H. pomatia* α-hemocyanin (30). It has been described as a classical core pentasaccharide (Man3-GlcNAc2) with the addition of a 1–6-linked fucose residue to the inner GlcNAc and a 1–2-linked xylose residue to the branching mannose residue. The unusual presence of xylose in an animal glycoprotein was later confirmed in several gastropods (Ref. 31 and references therein). Heavier glycoforms were also observed in the mass spectrum and could correspond to Gal, GalNAc, and MeGal-containing diantennary N-glycans that have also been characterized in *H. pomatia* α-hemocyanin (32).

**Binding Specificity Study**—Screening for oligosaccharide specificity (Fig. 2) confirmed the binding of HPA to glycans with terminal GalNAc residue, such as the blood group A and Forssman antigens and related oligosaccharides. HPA also appears to tolerate substitution on position 3 of GalNAc by a β-galactosidase residue as shown by its binding to Thomsen-Freidenreich (T) antigen and its sulfated or sialylated derivatives. Oligosaccharides containing a terminal β-GalNAc or α-GlcNAc residues are also recognized. These binding data confirm, although with some interesting variations (see “Discussion”), the observations made earlier (15, 33, 34). The only discrepancy is the binding of βGlcNAc1–4βGal1–4βGlcNAc trisaccharide. Because HPA does not bind any of the other βGlcNAc-topped oligosaccharides present in the microarray, this observation will need to be confirmed using a different methodological approach.

**Thermodynamics of the Interaction**—To characterize the interaction between HPA and GalNAc, the $K_d$ values and the thermodynamic binding parameters were determined using titration microcalorimetry, a method that is well suited to the characterization of protein-carbohydrate interactions (35). A typical titration curve for HPA binding to GalNAc is shown in Fig. 3. Using a mathematical model corresponding to a simple binding mode, the value of the dissociation constant was calculated to be $1.3 \times 10^{-4}$ M, which is in agreement with the modest affinity reported previously for blood group A pentasaccharide (33). The interaction is strongly driven by enthalpy because the free energy of binding (∆G value of $-22$ kJ/mol) has an enthalpic contribution, ∆H, of $-35$ kJ/mol counterbalanced by a strong unfavorable entropy contribution, T∆S, of $-13$ kJ/mol. Such an entropy barrier is often observed for the binding of carbohydrates to lectins (35).

**Overall Structure and Oligomerization**—Native HPA crystallized in lens-shaped crystals of space group P6$_3$2$_2$ ($a = b = 61.6$ Å, $c = 104.9$ Å) isomorphous to that previously described in a crystallization report (36). In the 2.5 Å resolution x-ray crystal structure, the asymmetric unit contains one monomer consisting of a six-stranded antiparallel β-sandwich (Fig. 4A). The β-strands are connected by short loops except for one longer
loop that connects strands A and B, forming a hairpin at one extremity of the β-sandwich. Strands A and E are connected by a disulfide bridge between cysteine residues 9 and 80. The refined electron density map suggested the presence of glycosylation at Asn-34 but was insufficient to be able to model the glycan, probably because of flexibility of the sugars.

The 3-fold crystallographic symmetry axis generates a tightly associated trimer (Fig. 4B). Strand C plays a special role in the trimerization because the first half (amino acids 45–50) participates in a strong hydrophobic cluster with the amino acids in the N- and C-terminal regions of the neighboring monomer, while its second half (52–56) creates a β-sheet with facing strand D (Fig. 5A).

The disulfide bridge between monomers is centered on a 2-fold axis with the involvement of two facing Cys-42 residues that are located on one loop of the monomer, close to both the N and C termini, creating a very elongated, tail-to-tail type of dimer (Fig. 5B). The resulting rod-shaped hexameric molecule (trimers of dimers) is shown in Fig. 4C.

The Monosaccharide Binding Site—Complexes obtained for HPA from co-crystallization with GalNAc produced prismatic crystals of space group H32 (a = b = 48.4 Å, c = 286.2 Å). Again, the asymmetric unit consists of one monomer and the hexamer is created by the 3-fold and 2-fold axes. The overall hexameric structure is very similar to that of native HPA, albeit with a few degrees difference in the orientation of one trimer.

FIGURE 6. Crystal structure of the complex between HPA and GalNAc. A, final ω-weighted 2F o − F c electron density map (blue contoured at 1σ) around the GalNAc residue. B, representation of the binding site with GalNAc (as sticks) and zinc (as ball) and hydrogen bonds represented by dashed lines. C, orthographic representation of the trimer.
Crystal Structure of Helix pomatia Lectin

relative to the other. In this high resolution structure (1.3 Å), electron density could be detected for one GalNAc residue in each monomer (Fig. 6A) in a binding site away from the dimerization interface. In addition, one Zn²⁺ cation is present close to the GalNAc binding site (Fig. 6B). This cation lies on a special position on the 2-fold axis of symmetry and is hexacoordinated by one nitrogen from the side chain of His-84, the two oxygen atoms of one acetate ion, and the symmetry equivalents of these three atoms. The presence of this cation occurred in only one commercial lot of protein and is likely due to an exposure of living snails to metal ions, because these animals have the capacity to accumulate metals (including zinc) in their midgut gland and other organs (37). The presence of the divalent cation is accidental but promoted intermolecular contacts that yielded a different space group and higher resolution.

The GalNAc binding site consists of a shallow pocket at the interface between two monomers (Fig. 6C) made up by part of the long loop connecting strands A and B, the extremities of strands D and E of the same monomer, and of strands C and F of the neighboring one. Eight hydrogen bonds are observed between the monosaccharide and the protein (Table 3. The axial O4 oxygen is the most buried and establishes three hydrogen bonds with the side chains of Arg-63, Trp-83, and Asp-55 from the neighboring monomer. Arg-63 is also hydrogen bonded to the GalNAc ring oxygen O5 and to O6. This O6 atom of the hydroxymethyl moiety is directed inside the site and interacts with the side chains of Asn-61 and Asp-55 of the second monomer. The oxygen O3 forms hydrogen bonds with the main chain oxygen of Gly-24 and bridges through a water molecule to the side chain of Arg-25. The carbonyl oxygen of the sugar N-acetyl group is hydrogen bonded to the Gly-24 main chain nitrogen. Interestingly, the methyl group of GalNAc is not involved in any hydrophobic contacts. Such contacts are rather limited because only the CH₂ group of C6 makes van der Waals contacts with Tyr-89 of the neighboring monomer and the CH group of C1 with His-84.

The reported higher affinity for αGalNAc can therefore be attributed to two main contributions. First, three hydrogen bonds are established with the axial O4 oxygen, therefore creating higher affinity for monosaccharides with the galacto configuration. Second, His-84 establishes a strong contact with C1 and sterically favors the occurrence of the α-anomer at this position.

N-Glycosylation Site—In the highest resolution crystal structure, i.e., the HPA-GalNAc complex, the refined electron density map confirmed the presence of glycosylation at Asn-34. Only the first GlcNAc could be unambiguously located. The density was not sufficient to be able to model the whole oligosaccharide, probably due to flexibility of the sugars located in the solvent channel between HPA hexamer.

Structural Similarities with Other Proteins—Although many lectins adopt a β-sandwich, or β-barrel fold, none of them has the same topology and binding site location as HPA. The β-sandwich fold with six strands arranged in a Greek key is shared by a large number of protein domains adopting the so-called immunoglobulin-like fold, and structural similarity searches result in a large list of crystal structures of unrelated function. Most of them include one or more additional strands. Disulfide bridges creating tail-to-tail dimers of β-sandwiches have been previously observed in HSP 90 co-chaperone p23 although with a different connectivity of the strands in each monomer (38).

The trimeric arrangement is reminiscent of that observed in another lectin involved in innate immunity, the fusoc binding agglutinin from Anguilla anguilla eel (AAA) (39). The β-barrel of AAA has a different fold than HPA, and the binding site also has an unrelated architecture and specificity. Nevertheless, the presentation of three binding sites on the flat surface formed by the association of the three monomers has strong similarities with distances between the sites going from 20 Å for HPA to 25 Å for AAA.

Occurrence of HPA-like Proteins—When searching databases, the peptide sequence that presents the highest identity with HPA (60%) is the recently characterized CH-III, the GalNAc binding lectin from the albumin gland of the garden snail Cepaea hortensis (40). HPA also has homology, albeit lower (54%), with a sequence deposited by the same group and described as H. pomatia lectin (Q575S3, HELPO). Because this latter sequence is much closer to the C. hortensis lectin sequence (81% identity), it would appear to be from a different snail subspecies than the one from which the commercial lectin is available. The three gastropod

<table>
<thead>
<tr>
<th>TABLE 3</th>
<th>Contacts between HPA and GalNAc</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GalNAc</strong></td>
<td><strong>HPA</strong></td>
</tr>
<tr>
<td>Hydrogen bonds</td>
<td></td>
</tr>
<tr>
<td>O7</td>
<td>Asp-26 N</td>
</tr>
<tr>
<td>O3</td>
<td>Gly-24 O</td>
</tr>
<tr>
<td></td>
<td>Wat-1*</td>
</tr>
<tr>
<td>O4</td>
<td>Trp-83 NE1</td>
</tr>
<tr>
<td>Asp-55 OD2*</td>
<td>2.67</td>
</tr>
<tr>
<td>Arg-63 NE</td>
<td>3.00</td>
</tr>
<tr>
<td>O5</td>
<td>Arg-63 NH₂</td>
</tr>
<tr>
<td>O6</td>
<td>Asn-61 ND1</td>
</tr>
<tr>
<td>Asp-55 OD1*</td>
<td>2.62</td>
</tr>
<tr>
<td>Arg-63 NH₂</td>
<td>3.24</td>
</tr>
</tbody>
</table>

Van der Waals

<table>
<thead>
<tr>
<th></th>
<th>Van der Waals</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>His-84</td>
</tr>
<tr>
<td>C6</td>
<td>Tyr-89b</td>
</tr>
</tbody>
</table>

**FIGURE 7.** Alignment of amino acid sequences. H. pomatia lectin with entry code Q575S3, CH-III from snail C. hortensis, SLL-2 isoforms from octocoral S. lochmodes, and discoids from D. discoides. The background is shaded according to similarity between the sequences. Structural information from HPA includes the extent of β-strands (arrows) and the amino acids involved in hydrogen bonds through their side chains (o) or backbone (x) atoms and those establishing hydrophobic contacts (+) to the ligand.
lectins have a similar size and are predicted to share the same hexameric arrangement because the cysteine residue involved in making the intermolecular disulfide bridge is conserved as displayed in the alignment of Fig. 7.

SLL-2, a lectin isolated from the octocoral Sinularia lochmodes, also displays significant similarity in both size (94 amino acids) and sequence (23–25%, depending on the isoforms) to HPA. This galactose-specific lectin is involved in the surface recognition of symbiotic dinoflagellate Symbiodinium, and the sequence has been shown to be related to the C-terminal region of discoids (41). These proteins, which mediate the intercellular cohesive properties of D. discoideum cells, are developmentally regulated lectins with a binding preference for GalNAc for discoid-1 and for Gal or GalNAc for discoidin-2 (42).

The sequence alignment of all of the HPA-like invertebrate lectins or lectin domains is shown in Fig. 7. The similarity is stronger in the C-terminal region, and all of the amino acids that are necessary for binding to O4 of galactose (Asp, Arg, and Trp) are strictly conserved. There are more variations in the region predicted to correspond to the first β-strand and the long loop. This loop, which in HPA is responsible for binding the N-acetyl group of GalNAc, has the same length in the other gastropod lectins and in discoidin-1, which all bind GalNAc, but is shorter in coral lectins and in discoidin-2, which are reported to bind preferentially to galactose.

CONCLUSION

Because of its specificity and multimeric molecular architecture, the HPA lectin appears to be a powerful weapon to agglutinate pathogens. Although the oligosaccharide target recognized by HPA on the surface of pathogens has not yet been identified, its specificity is close to that of tachylectin-2 that binds to GlcNAc and GalNAc and recognizes staphylococcal lipoteichoic acids and lipopolysaccharides from several Gram-negative bacteria (6). N-Acetyl groups are very frequently observed in active substances present on pathogen surfaces and are now referred to as pathogen-associated molecular patterns. The architecture of HPA brings two sets of three binding sites on opposite hydrophilic and charged faces of the molecule in an architecture that allows two bacteria to be bound at the same time, therefore promoting aggregation and efficient protection of the snail eggs.

HPA has been demonstrated to be a useful tool in histopathology (43). It is considered to bind GalNAc present on the Tn antigen, but not its galactosylated form, the T antigen. However, the glycan array data that we present here demonstrate that the T antigen could also be recognized. This is in agreement with a previous study (34). The glycan array data presented here should not be considered as quantitative and therefore do not allow one to determine range in affinity. Nevertheless, they indicate that binding specificities other than terminal oGalNAc should not be ruled out when analyzing HPA binding to complex glycans at the surface of cancer cells. Interestingly, the ganglioside GD2, a well known tumor-associated carbohydrate antigen (44), does also bind to HPA in our assay. This preliminary screening opens the route for finer characterization of carbohydrate antigens that are specifically bound by HPA on different tumor cell lines.

The use of HPA as cancer marker is limited due to its origin. The purification from animals as the only available source is a drawback for several reasons, the main one being the presence of impurities as demonstrated here with the occurrence of zinc in the commercial sample. The knowledge of both the cDNA sequence and the structural features now allows production of large quantities of recombinant HPA with constant quality. It also offers the possibility of lectin engineering to develop powerful new tools for cancer research.

HPA-like proteins were identified in three families of invertebrates where the lectins seem to play different roles: innate immunity in snail, symbiosis mediator in coral, and social life in slime mold. Sequence similarities have been pointed out between the coral lectins and some bacterial sequence regions (41). Indeed, using the HPA sequence as a probe, a search in microbial genomes indicates similarity with putative proteins from Silicibacter, Magnetococcus, and Rhodobacter. It is therefore likely that HPA-like domains will be present in both animals and bacteria. By analogy with the new fucose recognition fold that has been observed in A. anguilla agglutinin and in bacteria, and which is now referred to as F-type lectins (39), we herein propose the name of H-type lectin for the HPA-like lectins and lectin domains.

Acknowledgments—We thank the European Synchrotron Radiation Facility, Grenoble, for access to synchrotron data collection facilities. Stéphanie Boullanger (Centre de Recherche sur les Macromolécules Végétales) and Bernard Dublet (Institut de Biologie Structurale) are gratefully acknowledged for their expertise in mass spectroscopy.

REFERENCES

Crystal Structure of Helix pomatia Lectin

Biochemical and Structural Analysis of Helix pomatia Agglutinin: A HEXAMERIC LECTIN WITH A NOVEL FOLD
Jean-Frederic Sanchez, Julien Lescar, Valérie Chazalet, Aymeric Audfray, Jean Gagnon, Richard Alvarez, Christelle Breton, Anne Imberty and Edward P. Mitchell

doi: 10.1074/jbc.M603452200 originally published online May 16, 2006

Access the most updated version of this article at doi: 10.1074/jbc.M603452200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 41 references, 6 of which can be accessed free at
http://www.jbc.org/content/281/29/20171.full.html#ref-list-1