Glycobiology vol. 20 no. 12 pp. 1643–1653, 2010
doi:10.1093/glycob/cwq118
Advance Access publication on September 8, 2010

Characterization of the secondary binding sites of *Maclura pomifera* agglutinin by glycan array and crystallographic analyses

Jingwei Huang[3,4], Zan Xu[4], Die Wang[2,4], Craig M Ogata[5], Krzysztof Palczewski[3], Xavier Lee[4], and N Martin Young[1,6]

[3]Department of Pharmacology, School of Medicine, Case Western Reserve University, 2109 Adelbert Rd., Cleveland, OH 44106, USA; [4]Department of Cell Biology, Lerner Research Institute, The Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, OH 44195, USA; [5]GM/CA CAT Advanced Photon Source, Argonne National Laboratory, Argonne, IL 60439, USA; and [6]Institute for Biological Sciences, National Research Council of Canada, Ottawa, ON, Canada KIA 0R6

Received on May 26, 2010; revised on July 28, 2010; accepted on July 28, 2010

The *Maclura pomifera* agglutinin (MPA) recognizes the T-antigen disaccharide Galβ1,3GalNAc mainly through interaction of the α-GalNAc moiety with its primary site, but the interactions of the two flanking subsites A and B with aglycones and substituents other than Gal, respectively, are not well understood. We therefore characterized the specificity of MPA in more detail by glycan microarray analysis and determined the crystal structures of MPA without ligand and in complexes with Galβ1,3GalNAc and *p*-nitrophenyl α-GalNAc. In both sugar complexes, pairs of ligands created inter-tetramer hydrogen-bond bridging networks. While subsite A showed increased affinity for hydrophobic aglycones, it also accommodated several sugar substituents. Notably, a GalNAc-*O*-tripeptide, a Tn-antigen mimic, showed lower affinity than these compounds in surface plasmon resonance (SPR) experiments. The glycan array data that showed subsite B accepted compounds in which the O3 position of the GalNAc was substituted with various sugars other than Gal, but substitutions at O6 led to inactivity. Additions to the Gal moiety of the disaccharide also had only small effects on reactivity. These results are all compatible with the features seen in the crystal structures.

Keywords: glycan array/lectin subsites/*Maclura pomifera* agglutinin

Introduction

Lectins are usually classified according to their primary specificity for mono- or disaccharides (Goldstein and Hayes 1978). This approach reflects the fact that most lectins have a dominant site which accommodates one or two sugars, and from a practical point of view, this enables simple affinity chromatography procedures to be used for most lectins. Yet it is apparent that surrounding subsites can be important for the overall specificity of a lectin, an early example being the discovery of the role of a fucose residue in the recognition of glycopeptides by the mannose-specific lentil and pea lectins (Kornfeld et al. 1981). Exploration of lectin fine specificity, namely the roles of neighboring subsites, has historically been limited by the poor availability of the necessary libraries of carbohydrates. However, the creation of glycan microarrays that display several hundred carbohydrate compounds (Blixt et al. 2004) by the core facilities of the Consortium for Functional Glycomics (CFG) has now made it possible to define the specificities of lectins and other carbohydrate-binding proteins in far greater detail. In the present work, CFG glycan arrays (CFG 2010) were used to characterize the subsites of the plant lectin *Maclura pomifera* agglutinin (MPA).

We have previously reported the crystal structure of MPA (Lee et al. 1998). It recognizes the T antigen, Galβ1,3GalNAc, through its α-GalNAc residue, and it also binds α-Gal. Both MPA and its homolog jacalin, from *Artocarpus integrifolia*, are homotetrameric proteins whose monomers contain a 133-residue α subunit and a 20- or 21-residue β subunit, derived from the proteolytic cleavage of a propolypeptide (Young et al. 1991). These lectins and a lectin from *A. hirsuta* with 96% sequence identity to jacalin (Rao et al. 2004) have β-prism I structures, with monomer units formed from 12 β strands (Raval et al. 2004; Sharma et al. 2007). Two subfamilies of jacalin-like lectins are known, a Gal-specific one in which a key feature of the binding site is an N-terminal glycine residue formed by proteolytic cleavage of a precursor propolypeptide and a Man-specific family in which there is a loop structure at this point (Bourne et al. 1999; Raval et al. 2004; Rabijns et al. 2005).

Crystal structures of complexes of jacalin with a series of Gal and GalNAc ligands (Jeyaprakash et al. 2002, 2003) have been described in terms of three sites: the primary site recognizing the α-GalNAc moiety, a secondary site A for aglycones and a secondary site B for the Gal of the T antigen, or other substituents. This model also applies to MPA, where a structure with the T-antigen disaccharide showed that the Gal

[1]To whom correspondence should be addressed: Tel: +1-613-990-0855; Fax: +1-613-941-1327; e-mail: martin.young@nrc-cnrc.gc.ca
[2]Present address: Centre for Cancer Research, Monash Institute of Medical Research Medical Centre, Monash University, 27-31 Wright Street, Clayton, VIC 3168, Australia

moiety had only weak contacts with the protein but did form inter-tetramer bridges in the crystal (Lee et al. 1998). The aglycone site is dominated by the residue Tyr122 which in NMR studies appeared to adopt a different conformation when a sugar with an aglycone was bound (Weimar et al. 2000).

We present analyses of the binding properties of MPA as determined with the CFG glycan microarray, together with crystal structures of MPA–sugar complexes and related SPR data. Collectively, these data define the specificity and subsites of this lectin in much greater detail than was previously possible.

Results

Glycan array analyses

The binding properties of MPA were assessed in microarray and plate formats. The full data for both formats are in substantial agreement (CFG 2010); only the glycan microarray results are described here. The data are shown in Figure 1, and the most active compounds are listed in Table I. The compounds can be grouped into two subsets, with either GalNAcα or Galα at their nonreducing ends, respectively.

Table I. The most active ligands from the MPA microarray experiment

Glycan #	Compound	Fluorescence (%)[a]
161	GlcNAcβ1,3GalNAcα	100
88	GalNAcβ1,3GalNAcα	92.4
125	Galβ1,3GalNAcα	88.6
144	Galβ1,4GlcNAcβ1,3GalNAcα	87.6
158	GlcNAcβ1,2Galβ1,3GalNAcα	82.8
32	SO_3-3Galβ1,3GalNAcα	65.6
101	Galα1,3GalNAcα	64.0
202	Neu5Acα2,3Galβ1,3GalNAcα	63.8
163	GlcNAcβ1,3Galβ1,3GalNAcα	52.3
10	GalNAcα	43.5
112	Galα1,4GlcNAcβ	41.3
58	Fucα1,2Galβ1,3GalNAcα	35.2
7	Galα	29.0
113	Galα1,6Glcβ	17.8
214	Neu5Acα2,3GalNAcα	14.9

[a]Expressed as a percentage of the strongest ligand, whose fluorescence reading was 22,659.
[b]Compounds with a shorter linker, Sp0 compared to the Sp8 of the others (Blixt et al. 2004).

Ligand association constants

The binding characteristics of ligands suggested by the glycan microarray data and other ligands used in the crystallographic experiments were measured by SPR. The results are summarized in Table II. The model glycotripeptide had a significantly lower association constant for MPA than the T-antigen methyl α-glycoside (Me α-Galβ1,3GalNAc), while *p*-nitrophenyl α-GalNAc had an affinity more than 2-fold greater than that of methyl α-Gal, indicating the positive effect on binding of the *N*-acetyl function in the GalNAc compound is offset by negative effects from the *p*-nitrophenyl group, such as the repositioning of Tyr122 that we describe below. The affinity of MPA for the T-antigen methyl glycoside was approximately 10-fold weaker than that reported for jacalin, whose K_D was 1.2 μM (Jeyaprakash et al. 2003).

Crystal structures of native MPA and MPA complexes

To better define the binding properties of MPA, we structurally characterized the native protein and its complexes with *p*-nitrophenyl α-GalNAc and Galβ1,3GalNAc. The structure of the latter complex had previously been determined at a

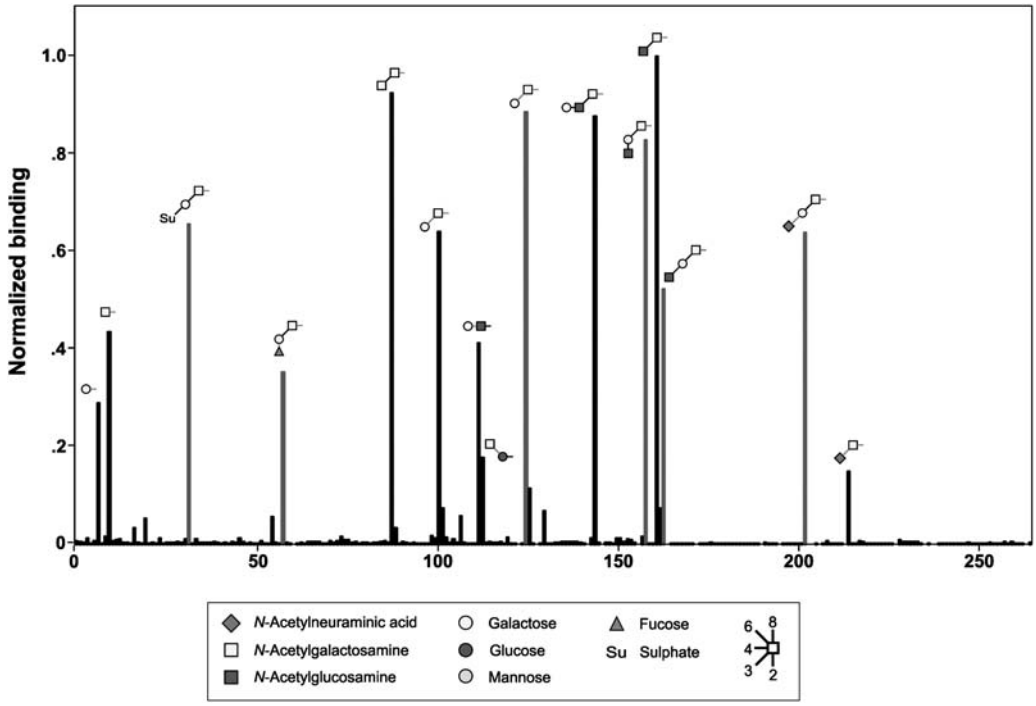


Fig. 1. Glycan array data for MPA. The blue bars indicate compounds related to Galβ1,3GalNAcα. Symbols for the sugars are given below, and α-glycosidic bonds are shown in red. Binding activities are scaled to that of the compound with the highest binding (fluorescence). The diagram is based on the CFG spreadsheets which can be accessed at http://www.functionalglycomics.org/glycomics/HServlet?operation=view&sideMenu=no&psId=primscreen_PA_v2_234_01202006.

Table II. Surface plasmon resonance measurements of ligand affinities

Ligand	K_D (μM)
Me α-Galβ1,3GalNAc	16
α-GalNAc-tripeptide	240
p-Nitrophenyl α-GalNAc	20
Me α-Gal[a]	52

[a]Determined by fluorescence spectroscopy (Young et al. 1989).

resolution of 2.2 Å (Lee et al. 1998). The crystallographic data are summarized in Table III.

The tetramer structure

A 1.55 Å resolution structure of the Galβ1,3GalNAc complex was refined using the 2.2 Å model (Lee et al. 1998); the root mean square (r.m.s.) deviation of the Cα atoms in the two structures was 0.218 Å. The differences were mainly in the side chain conformations, some of which (Lys45, Thr10, and Lys117) contribute to MPA tetramer formation or to crystal packing. The β subunit is better defined, showing that its N-terminal region clearly protrudes from the main body of MPA and interacts with another MPA monomer (Figure 2A). The interaction between the α and β subunits within an MPA monomer involved residues of the C-terminal half of the β subunit, whose peptide backbone is in juxtaposition with the last β strand (β11) of the α subunit (Figure 2B). Five main chain hydrogen bonds were identified for this interaction, which involves residues Phe127, Ile129, and Leu131 of β11 in the α subunit and Ile9, Val11, and Trp14 of β12 in the β subunit (Figure 2C).

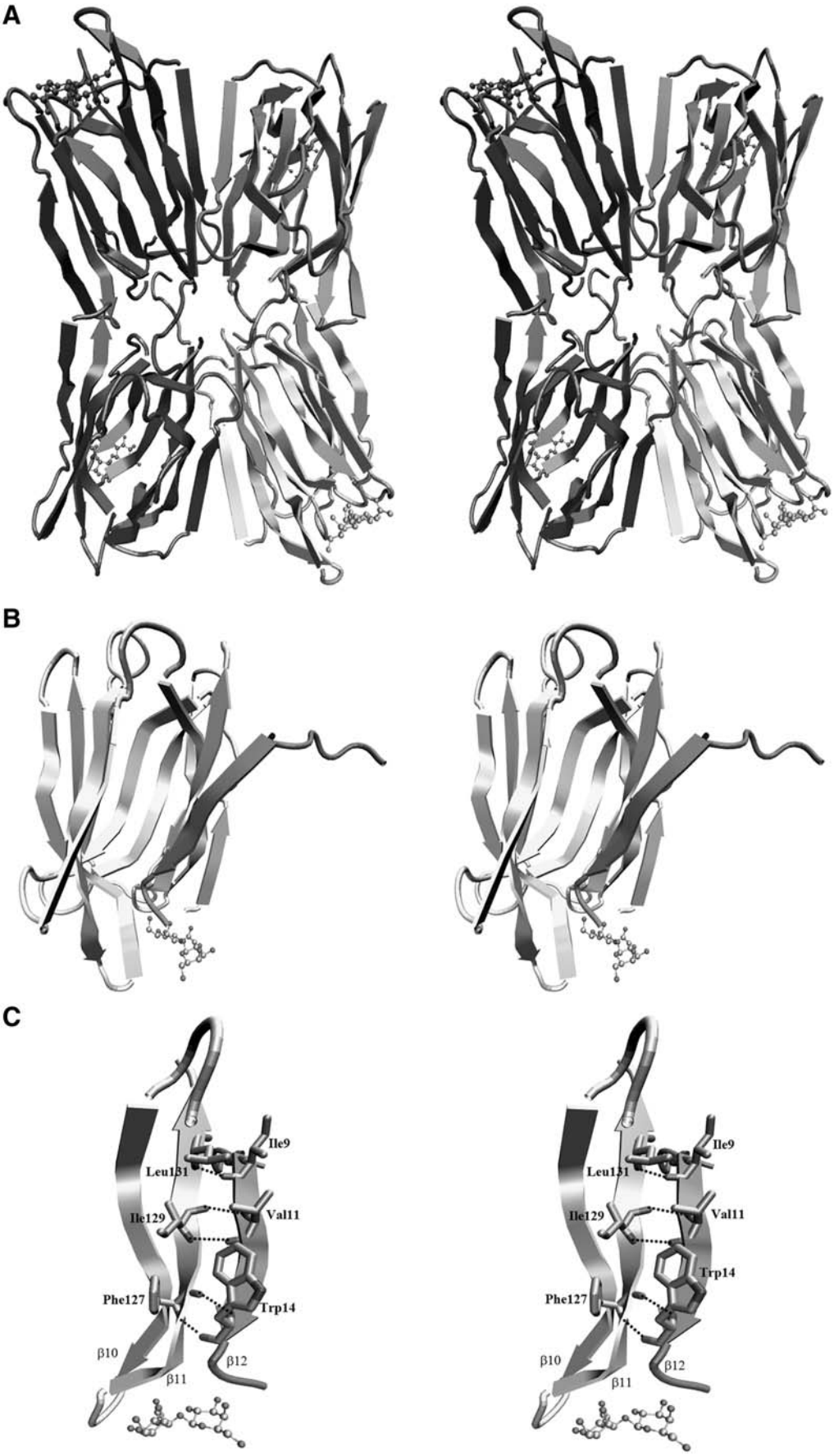
Binding of Galβ1,3GalNAc

The disaccharide associates with MPA in a region proximal to the β7 strand (residues 78–88), the loop connecting β10/β11 (residues 121–123), the β11 (residues 124–132), and to the N-terminal loop of the α chain (residues 1–2). There are two notable differences in the binding site in the higher resolution model. Firstly, the conformation of the Tyr122 side chain is different. This residue contributes to the affinity for T antigen by forming a hydrophobic environment along with other residues in the binding site. The torsion angles of the Tyr122 side chain are −85° for C-Cα-Cβ-Cγ and 55° for Cα-Cβ-Cγ-Cδ in the 1.55 Å structure, whereas they were −76° and −47° in the 2.2 Å resolution structure. Secondly, the O6 atom of the Gal showed a different orientation in the 1.55 Å resolution structure, namely a torsion angle of −88° for O6-C6-C5-C4 in the 1.55 Å model in contrast to 157° in the 2.2 Å model. There were also minor differences in the torsion angles of the interglycosidic linkage of the disaccharide, and more solvent molecules have

Table III. Summary of data collection and refinement statistics

	Galβ1,3GalNAc complex	*p*-Nitrophenyl αGalNAc complex	No ligand
Data collection			
Space group	$P6_422$	$P2_12_12_1$	$P6_422$
Unit cell parameters			
a, *b*, *c* in Å	66.54, 66.54, 147.95	69.09, 133.29, 200.05	61.75, 61.75, 149.53
α,β,γ in °	90, 90, 120	90, 90, 90	90, 90, 120
No. of observations	244,671	364,931	68,792
No. of unique observations	31,358	122,330	7981
d_{min} (Å)	1.5	1.9	2.25
Completeness (%) overall	98.2	83.8	92.4
Completeness (%) highest shell	86.6	46.6	95.1
R_{symm} (%)[a]	5.2	3.9	4.8
Refinement[b]			
Resolution range (Å)	20–1.55	20–2.1	20–2.25
No. of reflections			
Total	28,335	99,067	7756
Used for refinement	27,054 (σ_{cutoff} = 2)	95,957 (σ_{cutoff} = 2)	7536 (σ_{cutoff} = 2)
R_{free} calculation	2175 (7.5%)	9650 (8.9%)	632 (7.3%)
No. of non hydrogen atoms			
Protein	1151	9263	1168
Ligand	26	168	0
Water	155	880	72
R factor (%)	21.0	21.0	21.7
R_{free} (%)	22.7	24.4	26.3
Average temp. factor (Å^2)	19.54	27.8	29.8
r.m.s. deviation			
Bond lengths (Å)	0.005	0.007	0.006
Bond angles (°)	1.35	1.89	1.27
Ramachandran plot			
Residues in most favored regions (%)	88	84.6	85.7
Allowed regions	12	14.9	14.3
Generously allowed	0	0.4	0

[a]$R_{symm} = \Sigma_h\Sigma_i \mid I_{ih} - \langle I_h \rangle \mid / \Sigma_h\Sigma_i I_{ih}$, where I_{ih} is the *i*th observation of reflection *h* and $\langle I_h \rangle$ is the average intensity obtained from the same reflection observed *i* times.
[b]Statistical analysis of models was performed with the CNS suite, and the Ramachandran plot with Procheck in the CCP4 suite (Collaborative Computational Project, number 4 1994).



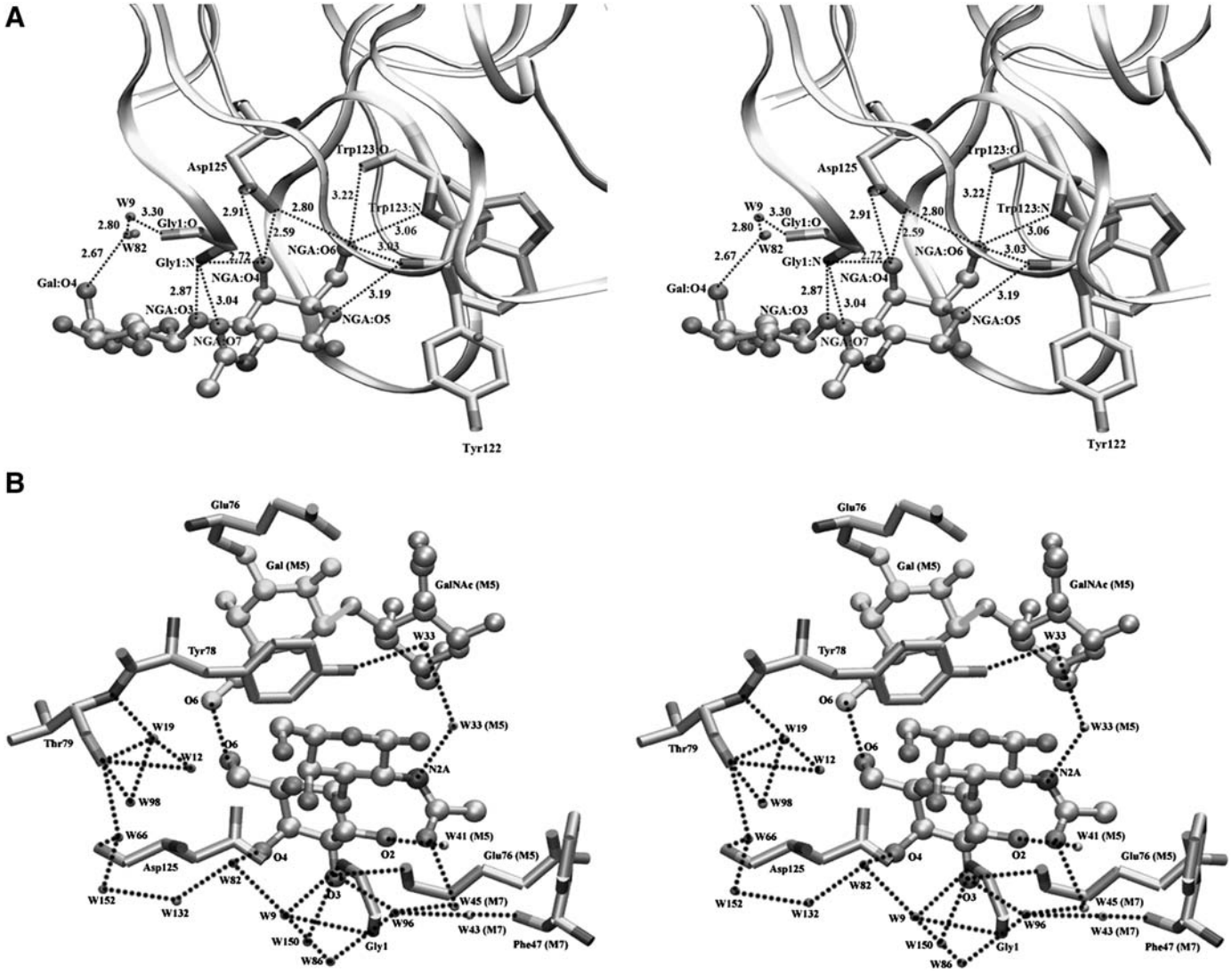


Fig. 3. Hydrogen-bonding diagrams of the disaccharide complex. (**A**) Hydrogen-bonding network of Galβ1,3GalNAcα with MPA. (**B**) Hydrogen bonding of the disaccharide with water and symmetrically related MPA monomers (M5 and M7).

been identified around the galactose residue in the higher resolution structure.

Consequently, the hydrogen bond network was revised from that reported previously (Lee et al. 1998). The GalNAc residue predominates in the binding by forming direct hydrogen bonds with Gly1, Tyr122, Trp123, and Asp125 of the α subunit (Figure 3A), while the Gal residue is surrounded by a number of ordered water molecules organized into a web of hydrogen bonds that involved residues belonging to three monomers from different tetramers (M1, M5, and M7, Figure 3B). There are no direct contacts from this subsite to the Gal. The only two nonwater hydrogen bonds around it are the O6 atoms forming a hydrogen bond directly with the O6 of galactose in a symmetric related disaccharide ligand (M5) and the O3 atoms forming a hydrogen bond with the peptide oxygen of Glu76 of M5 (Figure 3B). These two hydrogen bonds provide additional inter-tetramer association that aid in the crystal packing.

The inter-ligand and ligand–monomer cross-linkages are also revised. An inter-ligand hydrogen bond, O6–O6 of galactose residues from ligands of M1 and M5, was not seen in the previous 2.2 Å resolution structure. On the contrary, the indirect hydrogen bond between O6 of galactose and the peptide nitrogen of Thr79 via two water molecules (W171 and W157) proposed in the 2.2 Å resolution structure (Lee et al. 1998) is not observed in the 1.55 Å model, although two equivalent water molecules (W19 and W12) are present. These differences result from the change in the dihedral angle of O6-C6-C5-O5 in the Gal residue; O6 of the Gal in the 2.2 Å model pointed toward the prominent β7 and was closer to the Thr79, whereas it is further away from the ligand-binding pocket in the 1.55 Å structure.

Fig. 2. Structures of the MPA tetramer, monomer, and β chain region. (**A**) Crystal structure of the tetramer with the disaccharide Galβ1,3GalNAcα. (**B**) Monomer with the β chain. (**C**) Hydrogen bonding between α and β subunits in the MPA monomer. Note that the β numbers refer to the strands of the β structure, not to residues in the β subunit.

Table IV. Hydrogen bonds involved in ligand binding

Interaction[a]	Distance (Å)[b]	
	Galβ1, 3GalNAcα	*p*-Nitrophenyl αGalNAc
Hydrogen bonds between ligand and MPA		
GalNAc O5:N Tyr122	3.19	3.10
GalNAc O7:N Gly1	3.04	3.34
GalNAc O3:N Gly1	2.87	2.96
GalNAc O4:N Gly1	2.72	2.99
GalNAc O4:Oδ1 Asp125	2.59	2.60
GalNAc O4:Oδ1 Asp125	2.91	3.12
GalNAc O6:Oδ1 Asp125	2.80	2.78
GalNAc O6:N Tyr122	3.03	2.97
GalNAc O6:O Trp123	3.22	3.01
GalNAc O6:N Trp123	3.06	2.89
Gal O3:O Glu76 (M5)	3.27	
Water bridging between ligand and MPA		
GalNAc N2:O W33 (M5):O Tyr178 (M5)	2.95, 2.66	
Gal O2:O W41(M5):O Glu76 (M5)	2.74, 2.64	
Hydrogen bond between two ligands		
Gal O6:O6 Gal (M5)	2.30	

[a]M5 indicates a sugar or residue from a neighboring tetramer.
[b]Distance between the donor and acceptor.

More water molecules are located in the higher resolution structure, thus additional interactions between the ligand and MPA mediated by hydrogen bonds involving these new water molecules have been identified. For example, GalNAc N2 forms a hydrogen bond with a water molecule (a symmetrically related W33 [M5]) that also forms another hydrogen bond with the phenolic O of Tyr78 (M5). Gal O4 interacts with Gly1 O through two water molecules (W82 and W9). All the hydrogen bonds involved in the complex are listed in Table IV.

As mentioned above, the disaccharide is surrounded by residues with aromatic side chains, including Phe47, Tyr78, Tyr122, and Trp123 (Figure 4A), providing a hydrophobic environment to stabilize the sugar. The orientation of the phenolic ring of Tyr78 is parallel to the pyranose ring of the GalNAc at a distance of ~4 Å. On the other side, the same Tyr78 ring is parallel to the pyranose ring of the Gal of the neighboring symmetrically related disaccharide ligand at a distance of ~4.5 Å. Therefore, the van der Waals interactions resulting from hydrophobic ring stacking play an important role in both binding of the ligand to the lectin and the inter-tetramer interaction.

Structure of the complex with p-nitrophenyl α-GalNAc

The complex of MPA with *p*-nitrophenyl α-GalNAc had eight MPA monomers per asymmetrical unit, and the Fo-Fc map revealed clear densities for seven ligand molecules. There was no apparent difference between the polypeptide structures of the lectin associated with *p*-nitrophenyl α-GalNAc or with T antigen (r.m.s. deviation = 0.284 Å for Cα). Although the *p*-nitrophenyl α-GalNAc complex crystallized in a different space group than the T-antigen complex, the α/β subunit interactions and tetramer formation were identical. The different crystal packing is derived mainly from changes in the tetramer–tetramer interaction.

Final refinement of the complex structure resulted in excellent ligand map density. All the hydrogen bond interactions between the GalNAc residue and MPA found in the T-antigen complex were present, except that the distance between GalNAc O7 (the carbonyl oxygen) and Gly1 N is greater than optimal (Figure 4B and Table IV). The *p*-nitrophenyl substituent contributes additionally to the affinity of *p*-nitrophenyl α-GalNAc through interactions in the secondary site A. The *p*-nitrophenyl ring is parallel to the phenolic ring of Tyr122 at a distance of 3.8 Å. The stacking interaction between the two aromatic rings significantly stabilizes the complex. Furthermore, one nitro-group oxygen atom of the ligand forms a hydrogen bond with the side chain oxygen of Glu76. In comparison with the T-antigen complex, the position of the GalNAc was shifted slightly away from Gly1 and Tyr122. This results in an optimal stacking interaction between the two aromatic rings (Tyr122 and the ligand), but there is a weaker hydrogen bond interaction between GalNAc O7 and Gly1 N.

The structure of unliganded MPA

MPA crystallized in the same space group of $P6_422$ as the T-antigen complex. The 2.25Å structure was refined using the coordinates of the high-resolution MPA complex as a model. Comparisons of this structure with the ligand complexes revealed no major differences (r.m.s. deviations of Cα were 0.192 Å for the T-antigen complex and 0.336 Å for the *p*-nitrophenyl α-GalNAc complex). The hydrogen bonds and van der Waals forces mediating α/β subunit interaction or tetramer formation are also identical to those in the MPA complexes. Hence, the binding of the ligands to MPA does not induce any major conformational changes.

The conformation of the aromatic ring of Tyr122 in the native structure was notably different from that in the complex structure (Figure 4C). In native MPA, the torsion angles for C-Cα-Cβ-Cγ and Cα-Cβ-Cγ-Cδ of Tyr122 are −52° and 146°, respectively, deviating significantly from the Tyr122 residues in both the 1.55 Å and 2.2Å T-antigen structures. In the *p*-nitrophenyl α-GalNAc MPA complex, the angles were −87° and 41°, close to those of the T-antigen complex. Since it is perpendicular to the GalNAc pyranose plane of the incoming ligand, the orientation of the phenolic ring of Try122 in apo-MPA probably will not favor ligand binding, especially a ligand with a bulky aglycone such as *p*-nitrophenyl. Thus, when MPA is associated with a sugar ligand having an α substituent (methyl α-Galβ1,3GalNAc or *p*-nitrophenyl α-GalNAc), the conformation of the Tyr122 side chain is changed so that the ring plane becomes parallel to that of Tyr78, creating a cleft that allows optimal binding of the aglycone portion of the ligand. In the native state, the conformation of the side chain of Tyr122 appears to be stabilized by one water molecule (W8) bridging Tyr122 and Tyr78 via two hydrogen bonds, although the bond distances of these two hydrogen bonds are a little longer than optimal.

In native MPA, the binding site is occupied by three water molecules (W16, W3, and W6), which form hydrogen bonds with Gly1, Tyr122, and Asp125 (Figure 4C), residues that are all involved in GalNAc binding. The water molecule W16 forms a hydrogen bond with the peptide nitrogen of Tyr122, which is also the hydrogen bond donor for the oxygen of water molecule W3. Water W3 forms another hydrogen bond with the side chain oxygen (Oδ1) of Asp125. This Oδ1 oxygen also forms a second hydrogen bond with water W6, which also forms a hydrogen bond with the nitrogen of Gly1. In addition, these three waters were packed in close proximity, such that they could interact

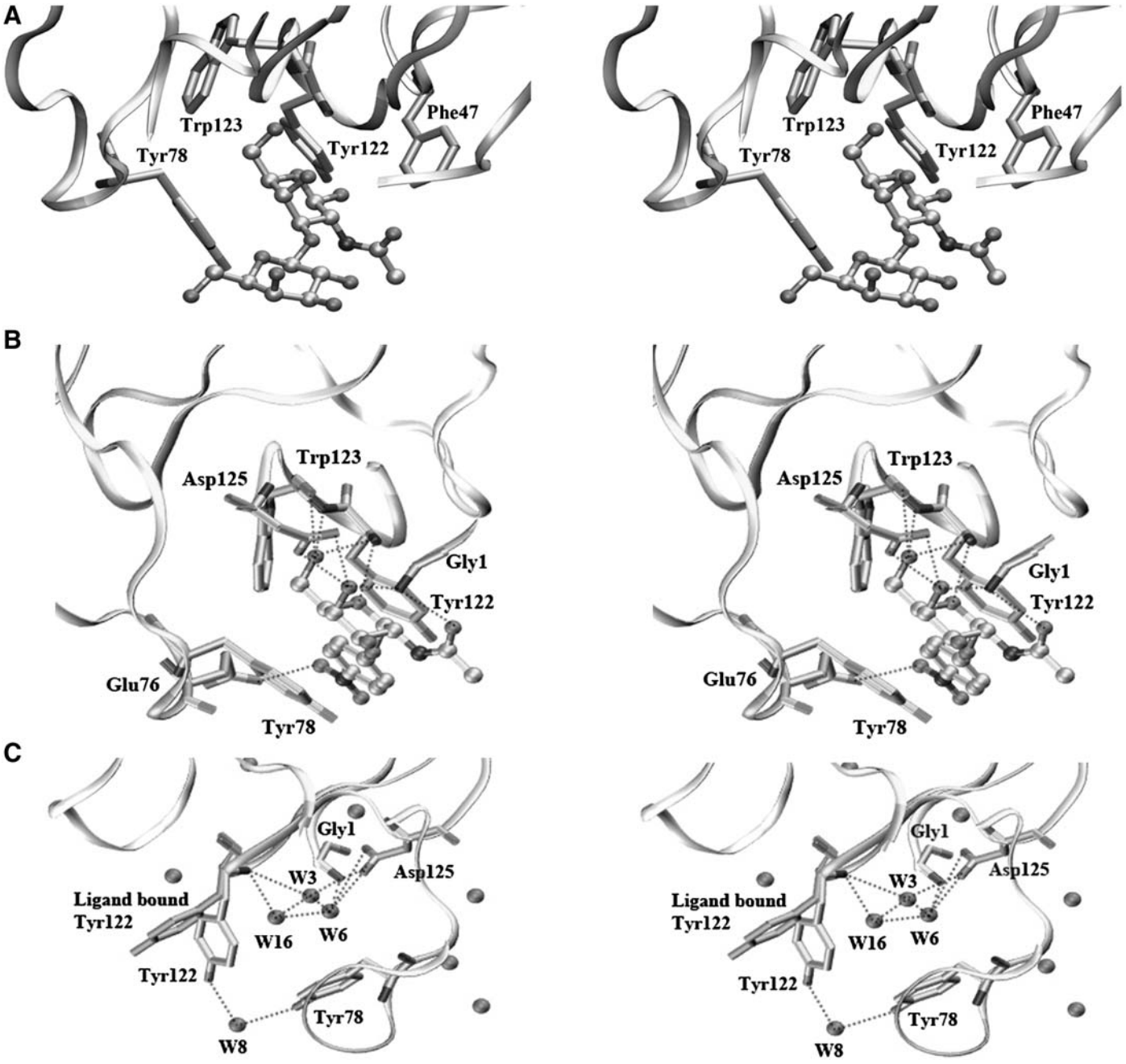


Fig. 4. Comparison of binding site regions. (**A**) The Galβ1,3GalNAcα complex. (**B**) The *p*-nitrophenyl α-GalNAc complex. (**C**) Apo-MPA. Three water molecules (W16, W3, and W6) occupy the binding site and form hydrogen bonds with Tyr122, Asn125, and Gly1 of the MPA α chain. The conformation of Tyr 122 from A is shown in green.

with each other via three hydrogen bonds, although the distances between W16 and W3, or W6 were marginally close for a hydrogen bond, 2.60 and 2.51 Å, respectively. The hydrogen bonds with Gly1, Tyr122, and Asp125 resemble the bonding pattern of the GalNAc moiety, with W16 being equivalent to GalNAc O5, W3 to O6 and W6 to O4.

Discussion

The utility of MPA and jacalin for *O*-glycan detection arises from their strong affinity for α-GalNAc, which is bound by the primary site of these lectins. However, the roles of the two secondary sites had not been clear. The microarray data substantially extend our knowledge of MPA's fine specificity, for which these subsites are chiefly responsible. The primary specificity for α-GalNAc and α-Gal is evident, and the overall array profile of MPA is similar to that of its homolog jacalin, which was determined from a commercial sample by the CFG Core H staff (CFG 2010). The jacalin data have higher error bars than the MPA data, but four compounds appear to be significantly recognized by jacalin, and not by MPA. These are GalNAcβ1,3Galα1,4Galβ1,4GlcNAcβ-Sp0 (compound #90), GalNAcβ1,3(Fucα1,2)Galβ-Sp8 (#89), Galβ-Sp8 (#17), and GalNAcα1,4(Fucα1,2)Galβ1,4GlcNAcβ-Sp8 (#87). These ex-

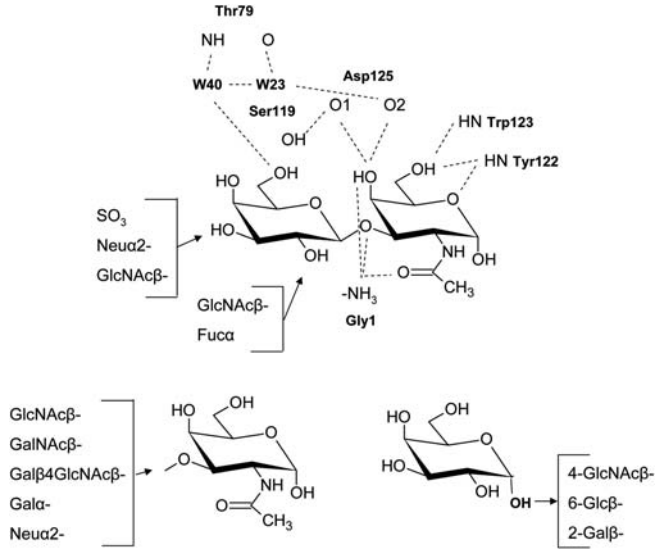


Fig. 5. Summary of the MPA ligands from the glycan microarray. The compounds are shown in three groups based on the T antigen (with its hydrogen bonds to the amino acid residues), α-GalNAc, and α-Gal derivatives, respectively.

ceptions do not have any structural element in common and are not attributable to the reported affinity of jacalin for mannose (Bourne et al. 2002; Jeyaprakash et al. 2005) or to contamination with the mannose-specific lectin artocarpin, also produced by *A. simplicifolia* (Barre et al. 2004). The data for MPA are summarized schematically in Figure 5, in three sets based respectively on α-GalNAc compounds, T-antigen homologs, and α-Gal compounds.

The fine discrimination among GalNAc compounds in the array data is remarkable, given the relatively small range of affinities found by SPR. This may be due to the polyvalent nature of the interaction of MPA with the glycan surfaces. This would magnify small differences in ligand affinities, and it has been shown that MPA has far higher avidities for polyvalent T- or Tn-antigen glycans such as mucins (Wu 2005). While the overall specificity of MPA for Gal and GalNAc would lead one to expect that it would recognize both blood group A and B structures, no such reactions were seen. The low reactivity of jacalin with A and B trisaccharides has been attributed to steric clashes between their Fuc moieties and the side chains of Tyr122 and Trp123 (Jeyaprakash et al. 2003).

The GalNAc binding site of MPA

The primary site of MPA drives the specificity of this protein, accommodating either α-Gal or α-GalNAc residues at terminal or internal positions in glycans. The strong specificity of MPA for these residues was confirmed by the glycan array experiments, which sampled a far greater number of ligands than had previously been available for testing.

The crystal structures described here show that the primary binding site of MPA is largely unchanged between the native and ligand-binding forms, and that there are only minor differences in H-bond lengths between the two ligand complexes (Table IV). The site's fully preformed nature leads to occupancy of those positions where the hydroxyl groups of the GalNAc ligand are bound by water molecules in the ligand-free structure. The same arrangement has been found in jacalin (Bourne et al. 2002).

The active compounds in the microarray and plate experiments all had α-GalNAc or α-Gal residues. Compounds that had substitutions at O6 (Figure 5) were inactive, in accordance with the major interactions of protein residues at this position in the crystal structure. Notably Neu5Acα2,6-GalNAc, the sialyl-Tn structure, was inactive as was reported for jacalin in frontal affinity chromatography (Tachibana et al. 2006). Similarly, the structure suggests substitutions at O4 would not be tolerated, but the repertoire of compounds on the array did not include any compounds of this type. Substitutions at O3 and O1 are discussed below.

The secondary subsite A

The behavior of the aglycone subsite is largely controlled by the key residue, Tyr122. Its role is reminiscent of tyrosine residues in antibody binding sites (Padlan 1990; Lo Conte et al. 1999) where the combination of its H-bonding ability with aromatic ring interactions, and low entropic loss due to conformational changes of the side chain, make Tyr122 uniquely capable of varied ligand interactions. Recently, an animal lectin, M-ficolin, was also shown to have a binding-site tyrosine whose conformation varied with different ligands (Garlatti et al. 2007). The different torsion angles, C-Cα-Cβ-Cγ and Cα-Cβ-Cγ-Cδ, of Tyr122 in native MPA and the MPA/T-antigen complex at 2.2 Å and 1.55 Å resolution, respectively, illustrate the flexibility of this side chain in response to ligand binding (Figure 4C). The flexibility of Tyr122 has also been demonstrated in jacalin T-antigen complexes (Jeyaprakash et al. 2002, 2003). Although the impact of Tyr122 ring orientation for the MPA Galβ1,3GalNAc complex may not be obvious, as it does not cause a spatial clash with the ligand, it does play an important role in the interaction with ligands having an α substituent, such as Me α-Galβ1,3GalNAc and *p*-nitrophenyl α-GalNAc. An NMR study of the Me T-antigen complex (Weimar et al. 2000) indicated a spatial conflict between the methyl group and the side chain of Tyr122, if its geometry was assumed to be that of MPA in the 2.2 Å model. This clash would be resolved if the torsion angle Cα-Cβ-Cγ-Cδ of Tyr122 is changed by 10°–30°, towards that in the jacalin Me α-Gal structure (Sankaranarayanan et al. 1996). In the 1.55 Å structure study of MPA with the T antigen, the Tyr122 torsion angles were −85° and 55°, and Tyr122 was nearly parallel to the ring of Tyr78, forming a cleft that would accommodate the methyl aglycone of Me α-Galβ1,3GalNAc.

In jacalin, the phenolic ring of Tyr122 forms a π...H–O interaction with OH and a π...H–C interaction with CH_3 of the GalNAc moiety in T antigens (Jeyaprakash et al. 2002, 2003). This specific π...H–O interaction is not observed in the MPA T-antigen complex structure, largely due to a different geometry of Tyr122. However, the significance of MPA Tyr122 in ligand association is demonstrated in the structure of MPA with *p*-nitrophenyl α-GalNAc. Here, the Tyr122 has a similar orientation to that found in the higher resolution structure of MPA with T antigen. The π–π stacking interaction between the aromatic rings of Tyr122 and the ligand explains the higher affinity for the *p*-nitrophenyl ligand compared to other ligands. This π–π interaction is so favored that the position of the GalNAc slightly

deviates from that seen in the T-antigen complex. This optimized interaction appears to cause a weakening of the hydrogen bond interactions between Gly1 N and the NAc O7, because this hydrogen bond distance (3.34 Å) is greater than the optimal range.

Unlike the favorable hydrophobic interactions occurring with the above aglycones, the tripeptide substituent provided a substantially weaker ligand. This was surprising, since the lectin and its jacalin homolog react well with glycoproteins that have *O*-linked glycans and can be used to purify them. In general, GalNAc compounds with one or more sugars at the reducing end were poor ligands, including compounds related to the blood group A antigen. However, compounds with terminal α-Gal groups were generally good ligands, including the strongest ligand in the plate format (Figure 1 and Table II). These included 6-Glc, 4-GlcNAc, and 2-Gal compounds. The lectin might distinguish between linear compounds and ones that have branched structures as in the tripeptide and blood group A cases. Overall, it can be seen that the secondary subsite A can accommodate various aglycones and sugars, which may utilize different conformations of Tyr122, and have binding energies above or below that of the T antigen.

The secondary subsite B

In the 2.2 Å structure of MPA and the T-antigen disaccharide (Lee et al. 1998) there were few interactions between the Gal of the ligand and the protein, but pairs of ligands formed bridges between tetramers in the crystal. The current structures show that *p*-nitrophenyl α-GalNAc ligands are also involved in tetramer–tetramer interactions by forming hydrogen bonds directly with the amino acid residues from adjacent tetramers. Hydrophobic interactions between the pyranose ring of Gal and the phenolic ring of Tyr78 also contribute to tetramer–tetramer interaction in the MPA T-antigen complex. The differences in the inter-tetramer interactions of the two ligands likely result in a difference in crystal packing as exemplified by the different space groups of crystals of the two complexes. However, the native MPA crystal had the same space group, $P6_422$, similar unit cell parameters and packed in the same way as the crystal of the disaccharide complex. This indicates that the inter-tetramer hydrogen bonds involving Lys45, Thr3, and Tyr78, and the van der Waals contact involving Phe47 and Gly1 are sufficient to mediate tetramer–tetramer interaction and crystal lattice formation.

The 1.55 Å structure reveals two major differences in the association of MPA with T antigen from the previous model (Figure 2). The orientation of the Gal O6 is changed, diminishing its interaction with Thr79 via hydrogen bonding through water molecules, and there is now an O6–O6 interaction between the ligand pairs. The Gal O6–Thr79 interaction, which is also observed in the jacalin T-antigen complex (Jeyaprakash et al. 2002), was believed to be responsible for the additional specificity and affinity of T-antigen binding by MPA (Lee et al. 1998). This interaction may play only a small role in T-antigen association with MPA, however, because the Me α-disaccharide is bound only ~3-fold more strongly than the Me α-Gal (Table II). The additional solvent molecules seen in the higher resolution structure make it clear that the Gal is stabilized in the crystal structure by hydrogen bonds involving multiple water molecules.

A wide range of ligand features, namely substituents on the T-antigen disaccharide or on the O3 of GalNAc, are tolerated at subsite A according to the microarray results (Figure 5). The Gal of the disaccharide can be replaced by several other sugars, including GlcNAc and GalNAc. The latter compounds had the highest affinities in the microarray assay, exceeding that of the T-antigen disaccharide itself, suggesting that the acetamido group may be able to form H bonds directly or through water molecules to the protein. A range of modifications of the Gal at O3 were tolerated, including NeuAc α2,3 and sulfate, as were two modifications at O2. These results are in accord with the exposure of these hydroxyls in the structure. The set of active compounds found in the array experiments differs from that recorded for peanut agglutinin (CFG 2010). This lectin is more specific for T antigen-related compounds and recognizes non-reducing terminal β-Gal residues, hence the Gal of the T antigen rather than the GalNAc predominates in its binding.

In conclusion, it is evident that MPA is capable of binding a wider range of α-GalNAc and α-Gal compounds than was previously realized. The plasticity of the aglycone subsite A and the exposed nature of the subsite B permit many substituents to occupy these sites. In contrast, substituents at the O6 position of the GalNAc, and probably O4, cannot be accommodated in the primary site. Therefore, the use of a single lectin as a reagent to detect a particular sugar, such as GalNAc, is clearly an unreliable strategy. Furthermore, the glycan array data delineate the fine specificities of many of the other Gal/GalNAc lectins that are available, hence much more information about the structure of a glycan can be obtained when two or more such lectins are used in concert.

Materials and methods

Lectins and ligands

MPA was purified as previously described (Young et al. 1989). Mono- and disaccharide ligands were obtained from Sigma-Aldrich Canada Ltd, Oakville, Ontario; the Tn-tripeptide Ala-(α-GalNAc-O)-Ser-AlaNH_2 was synthesized by J.-R. Barbier.

Glycan array assays

The glycan array data were collected by the CFG Core H staff using MPA labeled with Alexa Fluor 488 (Invitrogen, Canada Inc, Burlington, Ontario) and version 2 of the CFG glycan array; these data can be viewed in full on the CFG website (CFG 2010).

SPR

The interactions of sugar ligands with immobilized MPA were measured by SPR with a BIACORE3000 instrument (GE Healthcare, Bio-Sciences Inc, Baie d'Urfé, Quebec). Approximately 4500 resonance units (RUs) of MPA and 2000–4000 RUs of bovine serum albumin (Sigma-Aldrich Canada Ltd, Oakville, Ontario) as a reference protein were immobilized on research grade CM5 sensor chips (GE Healthcare), respectively. Immobilizations were carried out at protein concentrations of 50 μg/mL in 10 mM sodium acetate buffer pH 4.5, by means of amine coupling kits supplied by the manufacturer. In all instances, binding analyses were performed at 25°C in 10 mM HEPES buffer, pH 7.4 containing 150 mM NaCl and 0.005% surfactant P20, at a flow

rate of 20 μL/min. Between analyses, the surfaces were washed with running buffer only, i.e. without regeneration solution. The data were analyzed with the BIAevaluation 4.1 software (GE Healthcare).

Crystallization of MPA

The MPA complex was crystallized by the vapor diffusion method using sitting drops that contained either the disaccharide Galβ1,3GalNAc or *p*-nitrophenyl α-GalNAc. Only the α anomer of the disaccharide co-crystallized with MPA as reported earlier (Lee et al. 1998). The reservoir contained a mixture of 0.5 M Li_2SO_4, 12% polyethylene glycol 8000, and 1% octyl β-D-glucopyranoside buffered to pH 7.0 with 0.1 M HEPES. The protein concentration was 28 mg/mL. The stoichiometry of MPA to the ligand was 1:1.2. Native MPA was crystallized as described above, except no ligand was present.

Data collection and processing

Flash-frozen crystals of both native and complex forms were used for data collection on beamline X4A at the National Synchrotron Light Source, Brookhaven National Laboratory (Upton, New York). Because the crystals were previously determined to be in the high symmetric space group of $P6_422$ (Lee et al. 1998), a total of 90° with a 1° range for each oscillation was collected for native MPA. The high-resolution T-antigen complex data were collected in two passes, a fast low-resolution pass followed by a high-resolution pass with the detector moved to a closer distance. A total of 90° with a 0.6° oscillation range for each frame was collected for the MPA *p*-nitrophenyl α-GalNAc crystal in the orthorhombic space group $P2_12_12_1$. Data were indexed and scaled with the Denzo/SCALEPACK suite (Otwinowski and Minor 1997). Statistics of the data sets used for structure determination are presented in Table III.

Refinement of the native and complex structures of MPA

Because the space groups of the native MPA and the MPA T-antigen complex crystals were identical to those of the MPA structure at 2.2 Å (Lee et al. 1998), and the cell unit parameters are also very close, the coordinates of the model of the MPA complex at 2.2 Å were used directly to determine the high-resolution structure of the MPA complex with disaccharide, first at 2.2 Å and then extended to 1.55 Å. Structure refinement was performed with CNS (version 1.1) by first using rigid body refinement followed with energy minimized simulated annealing using maximally likelihood targets and B factor refinement (Kleywegt and Brunger 1996; Adams et al. 1997). Solvent molecules were picked from peaks of Fo-Fc maps according to the conditions for likely hydrogen bonding with neighboring atoms. The model was visually examined with the program O (version 7.01) (Jones et al. 1991). Manual correction was performed for the portion of the model that did not match the densities. The model was refined again with CNS. Omit maps were also calculated and used to reduce the model bias. Multiple cycles of CNS refinement followed by manual graphical interpretation with the program O were performed until no further improvement of the model was achieved. The model of the high-resolution MPA complex structure was then used to determine the native MPA structure. The same model was also used to determine the MPA *p*-nitrophenyl α-GalNAc complex, by molecular replacement using AMORE in the CCP4 package (Collaborative Computational Project, number 4 1994). Both refinements followed the same procedures described above. In the structure of the MPA T-antigen complex, the density of the first two residues at the N-terminus of the β-peptide was absent, but the main chain density was clearly visible in the difference Fourier map of native MPA. Thus, the coordinates for these two terminal amino acids could be obtained.

The MPA *p*-nitrophenyl α-GalNAc complex crystallized in space group $P2_12_12_1$. There were eight molecules in one asymmetric unit. A 4-fold noncrystallographic symmetry axis was used as a restraint during model refinement and released in the final cycles of the refinement. The final model resulted in clearly defined densities for seven *p*-nitrophenyl GalNAc ligands. The ligand density for the eighth MPA molecule was incomplete and therefore it is not included. The statistics of the final refined models of native MPA and the MPA complexes are summarized in Table III. Coordinates for the structures described in this paper have been deposited in the Protein Data Bank with codes 3LLZ, 3LM1, and 3LLY.

Graphic representation

Figures were generated either using the program O to produce plot files, then fed to the Mol-Ray (version 1.3) server (http://xray.bmc.uu.se/markh/), which is an interface to POV-Ray (http://www.povray.org) for ray tracing, or by Visual Molecular Dynamics (VMD 1.8.5).

Funding

This work was supported in part by the National Institutes of Health, Bethesda, MD (GM079191 to K.P.) and by the National Institute of General Medical Sciences of the NIH (GM62116 to the Consortium for Functional Glycomics). Use of the National Synchrotron Light Source, Brookhaven National Laboratory, was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences (contract no. DE-AC02-98CH10886). At the time of data collection, beamline X4A was supported by the Howard Hughes Medical Institute.

Acknowledgements

We thank Dr Richard Alvarez, University of Oklahoma, and Dr David F. Smith, Emory University, for the glycan array analyses at the CFG Core H facilities. We also thank Tomoko Hirama for the SPR measurements, Dr Robert Henning, Tom Devecseri and David C. Watson for excellent technical assistance, and Dr Warren Wakarchuk for helpful comments.

Conflict of interest statement

None declared.

Abbreviations

CFG, Consortium for Functional Glycomics; LacNAc, Galβ1,4GlcNAc; MPA, *Maclura pomifera* agglutinin; NMR, nuclear magnetic resonance; RU, resonance unit; r.m.s., root mean square; SPR, surface plasmon resonance.

References

Adams PD, Pannu NS, Read RJ, Brunger AT. 1997. Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc Natl Acad Sci USA.* 94:5018–5023.

Barre A, Peumans WJ, Rossignol M, Borderies G, Culerrier R, Van Damme EJM, Rougé P. 2004. Artocarpin is a polyspecific jacalin-related lectin with a monosaccharide preference for mannose. *Biochimie.* 86:685–691.

Blixt O, Head S, Mondala T, Scanlan C, Huflejt M, Alvarez R, Bryan MC, Fazio F, Calarese D, Stevens J, et al. 2004. Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc Natl Acad Sci USA.* 101:17033–17038.

Bourne Y, Zamboni V, Barre A, Peumans WJ, Van Damme EJM, Rougé P. 1999. *Helianthus tuberosus* lectin reveals a widespread scaffold for mannose-binding lectins. *Structure.* 7:1473–1482.

Bourne Y, Astoul H, Zamboni V, Peumans WJ, Menu-Bouaouiche L, Van Damme EJM, Barre A, Rougé P. 2002. Structural basis for the unusual carbohydrate-binding specificity of jacalin towards galactose and mannose. *Biochem J.* 364:173–180.

CFG. 2010. http://www.functionalglycomics.org/glycomics/publicdata/selectedScreens.jsp.

Collaborative Computational Project, number 4. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr.* D50:760–763.

Garlatti V, Martin L, Gout E, Reiser J-B, Fujita T, Arlaud GJ, Thielens NM, Gaboriaud C. 2007. Structural basis for innate immune sensing by M-ficolin and its control by a pH dependent conformational switch. *J Biol Chem.* 282:35814–35820.

Goldstein IJ, Hayes CE. 1978. The lectins: Carbohydrate-binding proteins of plants and animals. *Adv Carbohydr Chem Biochem.* 35:128–340.

Jeyaprakash AA, Rani GP, Reddy GB, Banumathi S, Betzel C, Sekar K, Surolia A, Vijayan M. 2002. Crystal structure of the jacalin-T-antigen complex and a comparative study of lectin-T-antigen complexes. *J Mol Biol.* 321: 637–645.

Jeyaprakash AA, Katiyar S, Swaminathan CP, Sekar K, Surolia A, Vijayan M. 2003. Structural basis of the carbohydrate specificities of jacalin: An X-ray and modeling study. *J Mol Biol.* 332:217–228.

Jeyaprakash AA, Jayashree G, Mahanta SK, Swaminathan CP, Sekar K, Surolia A, Vijayan M. 2005. Structural basis for the energetics of jacalin-sugar interactions: Promiscuity versus specificity. *J Mol Biol.* 347: 181–188.

Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A.* A47:110–119.

Kleywegt GJ, Brunger AT. 1996. Checking your imagination: Applications of the free R value. *Structure.* 4:897–904.

Kornfeld K, Reitman ML, Kornfeld R. 1981. The carbohydrate-binding specificity of pea and lentil lectins: Fucose is an important determinant. *J Biol Chem.* 256:6633–6640.

Lee X, Thompson A, Zhang Z, Ton-that H, Biesterfeldt J, Ogata C, Xu L, Johnston RA, Young NM. 1998. Structure of the complex of *Maclura pomifera* agglutinin and the T-antigen disaccharide, Galβ1, 3GalNAc. *J Biol Chem.* 273:6312–6318.

Lo Conte L, Chothia C, Janin J. 1999. The atomic structure of protein-protein recognition sites. *J Mol Biol.* 285:2177–2198.

Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276:307–326.

Padlan EA. 1990. On the nature of antibody combining sites: Unusual structural features that may confer on these sites an enhanced capacity for binding ligands. *Proteins.* 7:112–124.

Rabijns A, Barre A, Van Damme EJM, Peumans WJ, De Ranter CJ, Rougé P. 2005. Structural analysis of the jacalin-related lectin MornigaM from the black mulberry (*Morus nigra*) in complex with mannose. *FEBS J.* 272: 3725–3732.

Rao KN, Suresh CG, Katre UV, Gaikwad SM, Khan MI. 2004. Two orthorhombic crystal structures of a galactose-specific lectin from *Artocarpus hirsuta* in complex with methyl-α-D-galactose. *Acta Crystallogr D Biol Crystallogr.* D60:1404–1412.

Raval SR, Gowda SB, Singh DD, Chandra NR. 2004. A database analysis of jacalin-like lectins: Sequence-structure-function relationships. *Glycobiology.* 14:1247–1263.

Sankaranarayanan R, Sekar K, Banerjee R, Sharma V, Surolia A, Vijayan M. 1996. A novel mode of carbohydrate recognition in jacalin, a Moraceae plant lectin with a beta-prism fold. *Nat Struct Biol.* 3:596–603.

Sharma A, Chandran D, Singh DD, Vijayan M. 2007. Multiplicity of carbohydrate-binding sites in β-prism fold lectins: Occurrence and possible evolutionary implications. *J Biosci.* 32:1089–1110.

Tachibana K, Nakamura S, Wang H, Iwasaki H, Tachibana K, Maebara K, Cheng L, Hirabayashi J, Narimatsu H. 2006. Elucidation of binding specificity of jacalin toward O-glycosylated peptides: Quantitative analysis by frontal affinity chromatography. *Glycobiology.* 16:46–53.

Weimar T, Bukowski R, Young NM. 2000. The conformation of the T-antigen disaccharide bound to *Maclura pomifera* agglutinin in aqueous solution. *J Biol Chem.* 275:37006–37010.

Wu A. 2005. Polyvalent GalNAcα1→Ser/Thr (Tn) and Galβ1→3GalNAcα1→Ser/Thr (Tα) as the most potent recognition factors involved in *Maclura pomifera* agglutinin-glycan interactions. *J Biomed Sci.* 12:135–152.

Young NM, Johnston RA, Watson DC. 1991. The amino acid sequences of jacalin and the *Maclura pomifera* agglutinin. *FEBS Lett.* 282:382–384.

Young NM, Johnston RA, Szabo AG, Watson DC. 1989. Homology of the D-galactose-specific lectins from *Artocarpus integrifolia* and *Maclura pomifera* and the role of an unusual small polypeptide subunit. *Arch Biochem Biophys.* 270:596–603.