THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 273, No. 11, Issue of March 13, pp. 6312–6318, 1998
Printed in U.S.A.

Structure of the Complex of *Maclura pomifera* Agglutinin and the T-antigen Disaccharide, Galβ1,3GalNAc*

(Received for publication, October 22, 1997, and January 12, 1998)

Xavier Lee‡§, Andrew Thompson¶, Zhiming Zhang‡, Hoa Ton-that‡, John Biesterfeldt‡, Craig Ogata‖, Lulu Xu‡, Rosemary A. Z. Johnston, and N. Martin Young****

*From the ‡Department of Cancer Biology, Cleveland Clinic Research Institute, Cleveland, Ohio 44195, ¶European Synchrotron Radiation Facilities, Avenue des Martyrs, Grenoble, France 38042, ‖Howard Hughes Medical Institute, National Synchrotron Light Source, Brookhaven National Laboratory, Upton, New York 11973, and the **Institute for Biological Sciences, National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada*

Maclura pomifera **agglutinin is a tetrameric plant seed lectin with high affinity for the tumor-associated T-antigen disaccharide, Galβ1,3GalNAcα, and hence for many *O*-linked glycopeptide structures. Unlike members of most lectin families, it lacks both metal ions and Cys residues. The structure of its complex with Galβ1,3GalNAc was determined to 2.2 Å by first using multiwavelength anomalous diffraction with a lead derivative of the native protein, and then using molecular replacement with the unrefined structure as a model to solve the structure of the complex. The subunits share the β-prism architecture and three-fold pseudo-symmetry of the related lectin jacalin, with the 21-residue β-chains in the center of the tetramer. Interactions with the GalNAc predominate in the binding of the disaccharide. It forms a network of H-bonds with only one side chain, from an Asp residue, the amino group of the N-terminal Gly of the α-chain, and peptide backbone atoms of two aromatic residues. The Gal moiety does not H-bond directly with residues in the same monomer, *i.e.* there is no true subsite for it, but there are interactions through two water molecules. In the crystal, it interacts with residues in the binding site of an adjacent tetramer. The minimum energy conformation expected for the disaccharide is retained, despite its mediating the tetramer-tetramer interactions in the crystal packing. The resulting lattice is comparable to those seen for complexes of other lectins with branched glycopeptides.**

The disaccharide structure Galβ1,3GalNAcα occurs on the surface of tumor cells as a mucin-associated antigenic marker, termed the Thomson-Friedenreich or T-antigen. It is one of the few chemically well defined tumor antigens with a proven association with malignancy (1). The disaccharide is also the central element in *O*-linked glycopeptide structures on other mammalian glycoproteins. This structure can be recognized by several seed lectins from different plant families, including two that have been structurally characterized in complexes with the disaccharide, peanut agglutinin (2) and amaranthin (3). For a lectin to be useful as a detection agent for the T-antigen, it must show low affinity for possible cross-reactive structures on cells. These include the Galβ1,3GalNAcβ moiety of gangliosides, *N*-acetyllactosamine and sialylated forms of the T- and Tn-antigens. The peanut agglutinin has recently been engineered to improve its specificity (4).

Particularly high specificity for the T-antigen is shown by two lectins from the Moraceae plant family, jacalin (from *Artocarpus integrifolia*) and *Maclura pomifera* agglutinin (MPA).[1] They show low affinity for β-galactopyranosides (5, 6), including *N*-acetyllactosamine, and a >1000-fold preference for Galβ1,3GalNAcαMe (K_a for jacalin, 4×10^5 M^{-1}) compared with the β-glycoside (6). They also bind strongly the methyl and *p*-nitrophenyl α-glycosides of Gal and GalNAc, with K_a values of of $0.2–1 \times 10^5$ M^{-1} (5–7). MPA and jacalin are homotetrameric proteins comprising subunits with 133 residue α-chains and remarkably small β-chains of 20–21 residues (7–9). For jacalin, it has been shown that the chains arise from a precursor protein by cleavage of an N-terminal 39-residue piece and excision of a four-residue linker that joins the β-chain to the α-chain (10). These lectins show no sequence homology to any other plant lectin families, such as the legume lectins, wheat germ agglutinin, or ricin. Two other members of the MPA/jacalin family have recently been reported, one being a mannose-specific lectin also from *A. integrifolia* (11). The second is a lectin from a taxonomically distinct source, the rhizomes of *Calystegia sepium* (12), with specificity for mannose and maltose. Both these lectins are single-chain proteins.

Overall, MPA and jacalin are 85% homologous (8), but each displays both genetic and post-translational heterogeneity (9). Despite this, both have been crystallized without fractionation of the isoforms (13–14). However, MPA has fewer genetic forms, two (as compared with at least four for jacalin), no glycopeptide, and a dominant β-chain isoform (9). It therefore offers better prospects for structure determination at high resolution. The structure of jacalin complexed with the monosaccharide ligand methyl α-D-galactopyranoside has recently been reported (15). The basic architecture of the jacalin monomer units was a β-prism with internal three-fold pseudo-symmetry. We now report the structure of MPA complexed with the biologically more important ligand, the T-antigen disaccharide. This structure was solved by a combination of multiwavelength anomalous diffraction (MAD) and molecular replacement techniques, as heavy atom derivatives proved difficult to obtain.

EXPERIMENTAL PROCEDURES

Crystallization of MPA with Galβ1,3GalNAc—The complex was crystallized by the sitting drop method of vapor diffusion as reported previously (7). The data set used for the structure determination was

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (codes 1JOT and R1JOTSF) have been deposited in the Protein Data Bank, Brookhaven National Laboratory, Upton, NY.

§ To whom correspondence should be addressed. Tel.: 216-445-7270; Fax: 216-445-6269; E-mail: lee@xtal.ri.ccf.org.

[1] The abbreviations used are: MPA, *M. pomifera* agglutinin; MAD, multiwavelength anomalous diffraction.

TABLE I
Summary of data collection

	MPA + disaccharide	MPA + Pb			
Space group	$P6_422$	C222			
Unit cell (Å)	$a = b = 67.60$, $c = 149.26$	$a = 58.97$, $b = 119.50$, and $c = 155.41$			
Angles (°)	$\alpha = 90.0$, $\beta = 90.0$, $\gamma = 120.0$	$\alpha = 90.0$, $\beta = 90.0$, $\gamma = 90.0$			
d_{min} (Å)	2.2	2.9			
X-ray source	Rotating anode	Synchrotron			
Temperature	20 °C	100 K			
Wavelength (Å)		λ_1	λ_2	λ_3	λ_4
	1.543	0.9505	0.9472	0.9180	1.0056
No. of observations	98232	77934	77220	78922	78104
No. of unique reflections	10135	11980	11880	11958	12016
Completeness (%) overall	93.5	96.4	95.6	95.9	96.7
highest shell	58.3	90.5	90.2	90.3	90.7
Multiplicity	9.6	6.5	6.5	6.6	6.5
No. of unique Bijvoet pairs		9606	10097	10106	10216
R_{symm} (%)[a]	3.6	2.2	2.0	1.9	2.6

[a] $R_{symm} = \Sigma_h\Sigma_i \mid I_{ih} - <I_h> \mid / \Sigma_h\Sigma_i I_{ih}$, where I_{ih} is the ith observation of reflection h and $<I_h>$ is the average intensity obtained from the same reflection observed i times.

collected using a San Diego Multiwire type area detector (Table I).

Derivative Search—Soaking the above crystals with heavy atom compounds or co-crystallization all failed to yield heavy atom derivatives. Co-crystallization of the native protein with heavy atom compounds was then attempted. A new crystal form of C222 was obtained with *p*-hydroxymercurisulfonic acid. Hanging drops were used and the reservoir contained a mixture of 0.5 M Li_2SO_4, 12% polyethyleneglycol 8000, and 1% octyl-β-D-glucopyranoside buffered to pH 7.0 with 0.1 M Hepes. The concentration of the protein was 28 mg/ml, and it was mixed with twice the stoichiometric amount of *p*-hydroxymercuriphenylsulfonic acid. However, the phasing power of the mercury based on a MAD data set was very low, probably due to disorder of the mercury atoms in the crystals. These crystals were then used in soaking experiments, and a successful MAD candidate was obtained with 1 mM trimethyl-lead acetate solution (Table I). The crystals showed a weak absorption edge of lead when checked with synchrotron radiation in National Synchrotron Light Source, Brookhaven National Laboratory. They were flash-frozen in liquid nitrogen with buffer containing 15% ethylene glycol.

MAD Data Collection and Data Processing—A flash-frozen crystal was used for the data collection on BM14 at European Synchrotron Radiation Facilities (ESRF), Grenoble, France. Six data sets were collected at six wavelengths, three around the absorption edge of mercury and three at that of lead. These wavelengths were chosen to minimize the f' and maximize the f'' differences at the mercury and lead edges and to give both high energy and low energy "remote" wavelengths. No evidence of mercury anomalous scattering was seen, but the lead anomalous signal was very clear. A 1° range was used for each oscillation, and a total of 180° was collected at each wavelength. A 2.9-Å data set was collected using the ESRF Image Intensified CCD (II/CCD) detector (3.4-s readout time, 1242 × 1152 pixels). The data were corrected for spacial distortion and nonuniform response using the program FIT2D (16) and integrated using DENZO (17). Scaling was performed with SCALEPACK (17), keeping Bijvoet mates separate. The data were of very high quality with highest R_{symm} of 2.6% overall to 2.9 Å, and under 10% for the highest shells (3.0–2.9 Å). The R_{symm} clearly varied slightly with the presence of anomalous signal. The completeness and redundancy of this MAD data set were high (Table I), with the most complete data set having 96.7% completeness and the least 95.6%. An average redundancy is 6.5 (Bijvoet pairs included, *i.e.* a redundancy of 3, counting Bijvoets as separate).

MAD Phasing—The CCP4 program suite (18) was used for further processing; interwavelength scaling was performed with SCALEIT, and the anomalous and dispersive Pattersons were calculated with FFT. The anomalous Patterson maps were easily interpreted by hand calculation, yielding two lead sites, which were refined using the program VECREF (18). The dispersive Patterson showed exactly the same sites. The peaks corresponding to the two sites (Pb1 = 0.145, 0.214, and 0.217; Pb2 = 0.250, 0.180, 0.107) in the anomalous Patterson were 11.9 and 10.0 σ above the background. MLPHARE was used for the phasing using initial input f' and f'' values extracted from the measured absorption spectra using the programs CROMER,[2] XASFIT,[3] and KRAMIG.[3] The point of inflection was chosen as the reference wavelength. Phasing was attempted at 2.9-Å resolution and yielded an electron density map with a figure of merit of 73%, excellent phasing power, and low Cullis R factor (Table II). DM was used for solvent flattening and histogram matching to improve the electron density. The figure of merit after density modification improved to 84%. The map was displayed with the program O (19) and was of high quality and readily interpretable. A difference Fourier calculated using the phases obtained from the lead derivative did not display other significant sites.

Tracing of the Molecular Backbone—The program O (19) was used to trace the Cα backbone of the molecule with the command BATON. Since the map was of excellent quality and there is a single Trp in each of the two chains, tracing of the whole molecule was readily achieved. Two αβ subunits of MPA were found in the asymmetric unit. The side chains of the molecules were fitted in subsequently. The model used for the molecular replacement search was based on the better of the two molecules.

Molecular Replacement to Determine the MPA-Disaccharide Structure—Since the quality of the above preliminary model was exceptionally good when judged by its secondary structure, it was tried directly for molecular replacement to determine the structure of MPA complexed with the disaccharide. The program XPLOR (20) was used for a molecular replacement search, with only the 133 residues of the α subunit as model, and for subsequent refinement. The resolution range used for the search was 4.0–8.0 Å, and only reflections higher than 2 σ were used. The PC refinement method following a rotation search in X-PLOR was used to find the orientation of the search model. The orientation was $\theta_1 = 133.48°$, $\theta_2 = 2.41°$, and $\theta_3 = 280.54°$. It corresponded to the fourth highest peak in the rotation search. This peak was the highest peak in the PC refinement (height $5.9085^{-0.2}$) by only a small amount. However, the translation search provided confirmation as the maximum of the T-function was more than 12 σ high for the space group $P6_422$ as compared with 7 σ for $P6_222$. The R value was 50.91% *versus* 59.41%. Therefore the correct space group is $P6_422$. The fractional coordinates of the translation function are 0.137, 0.325, and 0.094. The height of the T-function also justified that this solution was a correct one.

Refinement of the Structure—The model was refined by simulated annealing with X-PLOR (20) (version 3.854) to 2.2-Å resolution using data with an F cutoff of 2 σ. As in the molecular replacement search, only the α subunit was used in the first two cycles. The first cycle started with rigid body refinement, followed by steps of slow cool refinement from 3000 to 330 K. The overall temperature used was kept at 15 Å^2. From the third cycle onward, 16 residues (positions 3–18) of the β subunit and the disaccharide were added in as the density became visible. Individual temperature factor refinement was included from the third cycle on, and manual correction of the model with graphics was performed between cycles. The R factor dropped to 0.229 after three cycles, with a free R factor of 0.291. At this stage, electron density for water molecules was clearly visible, and these were added with the program Waterpick (20) and the resolution range was changed to 6.0–2.2 Å. Water molecules with refined B factors above 60 Å^2 and density less than 1 σ in the $2F_o - F_c$ map were subsequently removed. The final model contains 91 solvent molecules. Only positional and individual temperature factor refinements were performed from the fifth cycle on. Details of the refinement statistics are given in Table III.

[2] D. T. Cromer, personal communication.
[3] W. Hendrickson and J. Smith, personal communication.

TABLE II
Phasing statistics from MPLHRE (18)

Data set	Overall		Overall phasing power		Overall R_{cullis} (ANO)[a]				
	R_{cullis} (CEN)	R_{cullis} (ACEN)	CEN	ACEN					
λ_1 (inflexion)					0.83				
λ_2 (peak)	0.67	0.56	1.33	2.44	0.84				
λ_3 (high remote)	0.67	0.54	1.35	2.53	0.81				
λ_4 (low remote)	0.68	0.55	1.34	2.51					
Resolution (Å)	10	7.3	5.9	4.9	4.1	3.6	3.2	2.9	Overall
FOM[b] centrics	0.91	0.86	0.83	0.82	0.80	0.86	0.84	0.77	0.83
FOM acentrics	0.49	0.69	0.79	0.76	0.73	0.74	0.74	0.67	0.72
FOM overall	0.64	0.74	0.80	0.77	0.74	0.75	0.75	0.68	0.73

[a] $R_{cullis} = \Sigma_{hkl} \| F_{PH} \pm F_P | - F_{H(calc)} | / \Sigma_{hkl} | F_{PH} \pm F_P |$ where F_P represents the reference or point of inflexion structure factors and F_{PH} one of the other three wavelengths.
[b] FOM, figure of merit.

RESULTS

Structure Determination—The complex of MPA and the T-antigen disaccharide Galβ1,3GalNAc was crystallized as reported previously (13), but the space group was redetermined as $P6_422$ or $P6_222$. There is one $\alpha\beta$ monomer in the asymmetric unit, and the V_m was 2.83 Å^3/Da. Extensive searches for heavy atom derivatives by soaking or cocrystallization with the complex were unsuccessful, possibly because of the low pH of the crystallization and because the primary sequence has no cysteine residues and only one histidine. In addition, MPA has no exchangeable Ca^{2+} or other metal ions. Co-crystallization trials of the ligand-free protein with heavy atoms led to a new crystal form in the space group C222 (Table I). Soaking experiments then produced a heavy atom derivative with trimethyl lead acetate, an excellent candidate for MAD. This diffracted beyond 1.7 Å and the lead absorption edge was detected at the National Synchrotron Light Source, Brookhaven National Laboratory. MAD was therefore used to determine the MPA structure based on this single derivative. Data sets for four different wavelengths in the vicinity of the lead absorption edge were collected at the ESRF, Grenoble, using an Image Intensified CCD detector (Table I). Two lead sites were easily identified in the anomalous and dispersive Patterson maps, and a high quality electron density map was produced based on the MAD phasing. A model of MPA was built from this map and it was used without refinement to determine the structure of the complex of MPA with the disaccharide by molecular replacement. The model of the complex was refined with the program X-PLOR, and the statistics for the final model are given in Table III. The sequences used for the model were those of the dominant α-chain and the β_{2a} isoform (8, 9). The root mean square deviation between the initial MAD model of the α-chain and the refined structure was only 0.8 Å. An $F_o - F_c$ "omit" map of the disaccharide is shown in Fig. 1. The lead atoms in the ligand-free crystal are located in the disaccharide binding site, interacting with the amino group of the terminal Gly. This offers a further explanation for the difficulties in obtaining heavy atom derivatives. Its location also prevents meaningful comparison of the native and ligand-filled binding sites.

Oligomeric Structure and Crystal Packing—Fig. 2 shows the crystal packing in the unit cell. There is one $\alpha\beta$ subunit per asymmetric unit and 12 ($\alpha\beta$) monomers in the unit cell. The contacts between the tetramers are in the region of the T-antigen binding site, and include intermolecular disaccharide-protein contacts. Fig. 3 shows the overall structure of the MPA tetramer and T-antigen complex.

Structure of the MPA Monomer—The overall MPA structure is very similar to that of jacalin in its folding. As predicted from the CD spectra (7) and sequence (8), MPA is a β-sheet protein. In the refined model of the monomer, there are 12 β-strands that form three subdomains (Fig. 4). The four β-strands of each subdomain form a Greek key motif, *i.e.* two anti-parallel pairs.

TABLE III
Summary of refinement statistics of final model

Resolution range (Å)	6.0–2.2
No. of reflections	
Total	10135
Used for refinement (σ_{cutoff} = 2)	9636
R_{free} calculation	481
No. of non hydrogen atoms	
Protein	1166
Sugar	26
Water	91
R factor (%)	17.1
R_{free} (%)	23.1
Average of the temperature factor (Å^2)	
α-Chain	
Main chain	20.3
Side chains	24.1
β-chain	
Main chain	28.5
Side chains	35.0
Ligand	27.5
Solvent molecules	44.2
Root mean square deviations from ideality	
Bond lengths (Å)	0.006
Bond angles (°)	1.302
Improper (°)	0.656
Ramachandran plot	
Residues in most favoured regions (%)	84.0
Residues in additional allowed regions (%)	16.0
Overall G factor[a]	0.10

[a] From the program PROCHECK (24).

The overall structure is a three-fold pseudo-symmetric β-prism. Interestingly, the three-fold nature was also predicted from the protein sequencing based on a possible internal repeat in jacalin (8). The β-prism's core is stabilized partly by hydrophobic interactions and partly by hydrogen bonding, including interactions of five water molecules deeply buried among the three subdomains. The interaction between the α- and β-chains is through six interstrand hydrogen bonds between the β-chain and the last segment of the α-chain, from residue 125 to residue 131. Together, these strands form half of the third Greek key subdomain. No electron density corresponding to the first two residues of the β-chain was visible. There was electron density for the last two residues of the β-chain, but it was too weak to uniquely locate the side chain atoms. Therefore, these four residues were not included in the final model. The N terminus of the β-chain projects out of the globular structure of the $\alpha\beta$ monomer into the center of the tetramer (Fig. 3).

Interaction of MPA with T-antigen—The connecting loops between β-strands in the distal region of the monomer plus the N-terminal residue of the α-chain form the ligand-binding site (Fig. 4). The GalNAc moiety of the disaccharide fits in a groove-shaped subsite formed by the N-terminal residue (Gly-1), and the turns of residues 46–49, 76–82, and 122–125. There are no water molecules directly involved in the GalNAc-protein interactions. Fig. 5 shows the interactions of the GalNAc with some

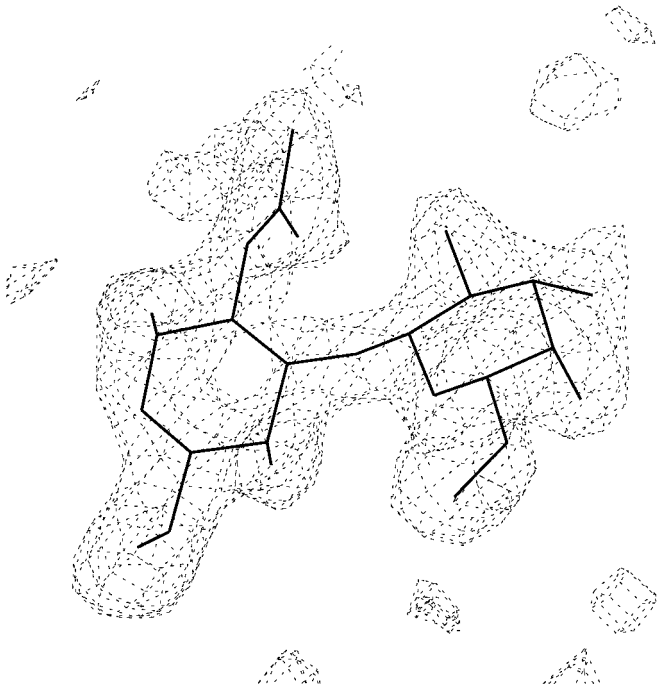

FIG. 1. **Quality of the final electron density map in the disaccharide region.** The $F_o - F_c$ omit map is contoured at 2.5 σ. The *dashed lines* show the experimental electron density in the binding site, and the disaccharide structure is shown by *solid lines*.

of the MPA residues. This sugar has the expected chair conformation with one side facing the amino group of Gly-1, Tyr-122, Trp-123, and Asp-125. Its hydrogen bonding interactions are summarized in Table IV, and further details are given in Fig. 6. The O-3, O-4, and the carbonyl oxygen of the *N*-acetyl group (O-7) all form hydrogen bonds with the Gly-1 amino group. The O-4 also hydrogen bonds to both carboxylate oxygens of Asp-125. The ring oxygen O-5 forms an H-bond with the peptide nitrogen of Tyr-122. The O-6 is at the bottom of the pocket in a network of hydrogen bonds with the Tyr-122 peptide nitrogen, the peptide nitrogen of the Trp-123, and one carboxylate oxygen of the Asp-125 side chain, which are in a tetrahedral array with C-6. The carbonyl oxygen of Trp-123 is slightly beyond H-bonding distance of the O-6 at 3.35 Å, and interacts with the peptide nitrogen of Gly-121 (3.10-Å). The carboxylate oxygens of Asp-125 also interact with a water molecule, W-23, and the hydroxyl of Ser-119. The distance between the Gly amino group and the carboxylate OD_2 of Asp-125 is 3.44 Å (Fig. 6). On its other face, the GalNAc is surrounded by the side chains of Phe-48, Tyr-78, Tyr-122, and Trp-123 with the phenol ring of the Tyr-78 parallel to the GalNAc ring (Fig. 5). The distance between this phenol ring and the GalNAc ring is approximately 4.0 Å. These hydrophobic interactions contribute significantly to the overall interaction between GalNAc and MPA.

The Gal moiety of the disaccharide protrudes from the end of the GalNAc pocket and is located between symmetrically related tetramers in the crystal (Fig. 7). There are no direct H-bonds to the protein, but O-6 interacts with a water molecule W-40, which in turn H-bonds to the peptide NH of Thr-79 and to another water molecule, W-23. The latter water molecule in turn H-bonds to the OD1 of Asp-125. Although there is no subsite for the Gal, its interaction with the protein through these water molecules accounts for the higher affinity of the disaccharide compared with monosaccharides. The Gal is in close contact with the neighboring tetramer through its C-2 and C-3 hydroxyls. Its O-3 forms a hydrogen bond with the Glu-76 peptide oxygen (3.15 Å) of the symmetrically related molecule, whereas water molecule W-59 forms hydrogen bonds with O-2 (2.82 Å) and the same Glu-76 peptide oxygen (3.08 Å). The Gal ring is also parallel to the phenol ring of Tyr-78 of the symmetrically related molecule at a distance of approximately 4.0 Å. There is also one protein-protein interaction, a H-bond between the two Tyr-78 hydroxyls (3.18 Å).

The overall conformation of the disaccharide is within the minimum energy area calculated for the disaccharide with the GalNAc as the α anomer.[4] The interglycosidic angles are ϕ 80° and ψ 128°. This suggests that the intertetramer interactions between the galactoses in the crystal do not alter the disaccharide conformation from its bound state with MPA in solution.

DISCUSSION

Structure and Processing of MPA—As expected from their sequence homology, MPA and jacalin have a common tertiary and quaternary structure. All of the MPA α-chain residues could be positioned in the electron density, and all except the two N-terminal and the two C-terminal residues of the β-chain. In the jacalin case, the three N-terminal and two C-terminal residues of the β-chain could not be placed. The jacalin β-chain is one residue shorter as well, so two more residues were seen in the MPA β-chain, *i.e.* 17 of 21 residues compared with 15 of 20 for jacalin. The C-terminal end is created by the post-translational proteolytic cleavage that removes four linking residues between the β- and α-chains in the jacalin precursor (10), so its mobility in MPA is not surprising. The jacalin processing also removes 39 residues that precede the β-chain, and the rather hollow center of the tetramer occupied only by the four crossing β-chains suggests these precursor segments could fit within the core region. The processing also creates a key ligand binding group, the amino group of the terminal Gly residue of the α-chain. However, whether the precursor protein has activity, *i.e.* if the post-translational proteolytic processing is essential for activity, cannot be predicted with certainty from the importance of this $-NH_2$ group. The recently described MPA homologs, artocarpin and Calsepin, that bind mannose and maltose do not undergo internal cleavage (11, 12) and Calsepin has a deletion at the position of the four linker residues in jacalin.

Carbohydrate Specificity of Moraceae Lectins—The overall H-bonding scheme for the MPA-disaccharide complex (Fig. 6) shows that the ligand is bound by MPA predominantly through the GalNAc unit, both through hydrogen bonds, mainly to atoms in the peptide backbone, and van der Waals interactions with a cluster of Tyr, Phe, and Trp residues. These interactions mainly account for the strong affinity and specificity of MPA for the T-antigen disaccharide. In contrast, the Gal projects from the protein and H-bonds only through two water molecules, while also interacting with the neighboring tetramer in the crystal. The dominance of the GalNAc in the overall recognition of the disaccharide is in marked contrast to the interactions found between peanut agglutinin and the same disaccharide (2). In the latter structure, the Gal O-3, O-4, and O-6 and the GalNAc O-4 all form H-bonds to side chains or Gly backbone atoms, plus there are a series of H-bonds through two pairs of water molecules with Gal O-2 and the carbonyl oxygen of the GalNAc, respectively. The strength of the disaccharide recognition by MPA despite its fewer bonding partners may arise from so many of the H-bonds being to charged groups, *i.e.* the Asp-125 carboxylate side chain and the terminal $-NH_3^+$ of Gly-1. Such H-bonds have been shown to be stronger than ones with neutral partners (27). The van der Waals interactions of the sugars with aromatic residues will also contribute not only to the binding of the disaccharide to the site but also to the

[4] J.-R. Brisson, personal communication.

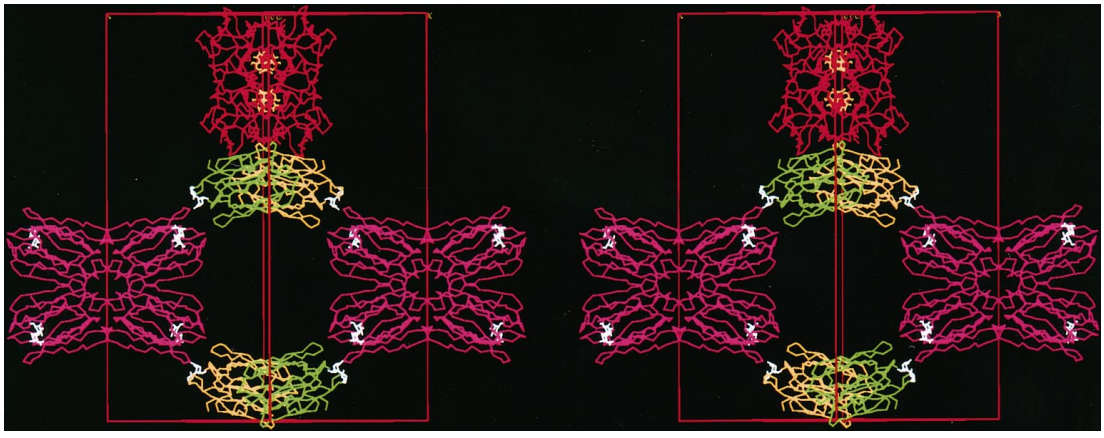

FIG. 2. **Crystal packing of the MPA-tetramer-disaccharide complexes.** One unit cell is shown edge on in stereo, with the *x* and *y* axes in the plane and the *z* axis vertical. The unit cell contains two half-tetramers (*pink*) and monomers from eight other tetramers (two *green*, two *yellow*, and four *red*). The disaccharide ligands (*white* or *yellow*) are at the corners of the tetramers and mediate the contacts between them. These interactions result in the formation of a lattice-like structure, as described under "Discussion."

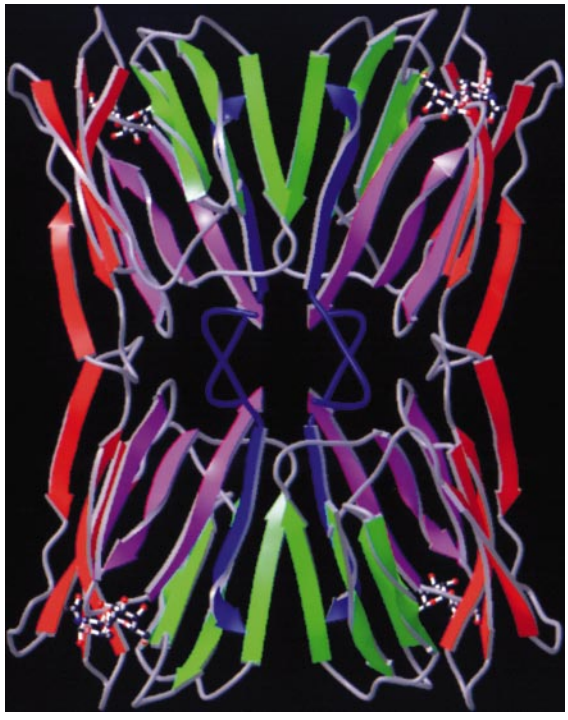

FIG. 3. **Backbone structure of the tetramer with the bound ligand.** The Greek key subdomains of each monomer are shown in *red*, *green*, and *purple* with the β-chain in *blue* (see Fig. 4). This figure and Figs. 4 and 5 were prepared with MOLSCRIPT (21) and RASTER 3D (22).

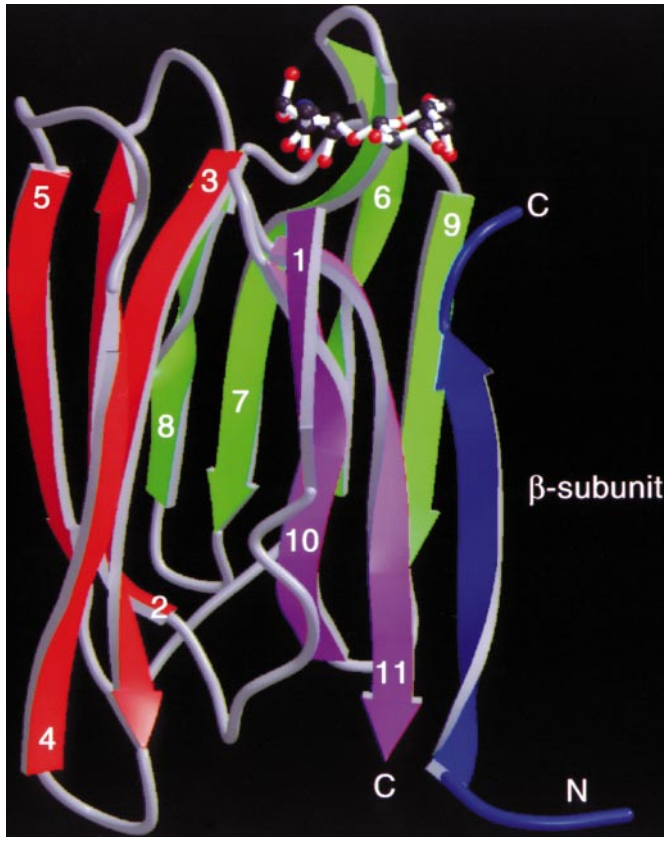

FIG. 4. **Backbone structure of the αβ monomer with the ligand.** Three subdomains together form the faces of a β-prism structure for the αβ monomer. The β-strands of the Greek key subdomains are colored as follows. Strands 2–5 form the first subdomain in *red*, strands 6–9 form the second subdomain in *green*, and strands 1, 10, and 11 in *purple* plus part of the β-chain in *blue* form the third subdomain.

intertetramer interactions of the Gal.

The structure satisfactorily explains the specificity characteristics of MPA. Although the Galβ1,3GalNAc ligand would be present as a mixture of α- and β-anomers, only the α-anomer is bound. The positioning of the GalNAc against the protein surface would make it impossible for any β glycoside of GalNAc or Gal to bind in the same orientation, explaining the strong anomeric preference of the Moraceae lectins (5, 6). The preference of MPA for the T-antigen compared with the Galβ1,3GalNAcβ element of gangliosides and lactosamine structures is also attributable to this feature. However, there is clearly an open area above the GalNAc O-1 for an α-linked aglycone (Fig. 5). NMR studies of an *O*-linked glycopeptide (25) suggest that the disaccharide when *O*-linked to Ser or Thr will be perpendicular to the peptide backbone. This arrangement would be easily accommodated in MPA. The higher affinity for α-nitrophenyl glycosides can be attributed to van der Waals interactions of the nitrophenyl ring with the nearby aromatic side chains. For the Gal, several of its hydroxyls are exposed and apparently could be substituted without impairing binding to MPA. There have been reports that jacalin can accept the 2,3′-sialylated glycopeptide (26). The natural ligand for the Moraceae lectins is of course unknown, but the binding site should be capable of accepting a polysaccharide with a Galβ1,3GalNAcα internal element as readily as it does the short *O*-linked glycopeptides.

The GalNAc moiety of the Galβ1,3GalNAc disaccharide is

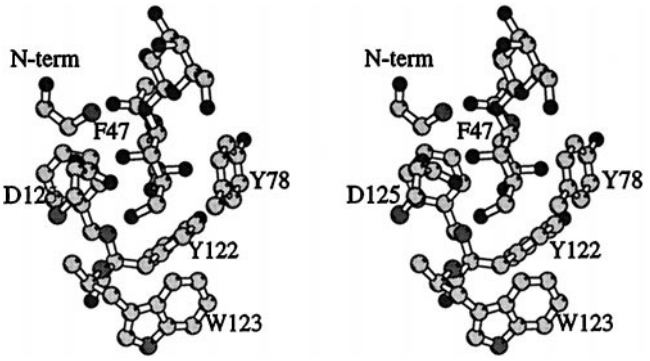

FIG. 5. **Binding of the GalNAc moiety to MPA.** H-bonds from backbone and side chain atoms and aromatic interactions in a pocket around the GalNAc contribute to the strong interaction between the GalNAc moiety of the disaccharide and the MPA binding site.

TABLE IV
Hydrogen bonds between ligand atoms and MPA residues

Ligand atom	Protein atom	Distance (Å)
GalNAc O3	Gly-1 NH	2.70
GalNAc O4	Gly-1 NH	2.82
GalNAc O4	Asp-125 OD1	2.81
GalNAc O4	Asp-125 OD2	2.85
GalNAc O5	Tyr-122 NH	3.15
GalNAc O6	Tyr-122 NH	3.08
GalNAc O6	Trp-123 NH	2.98
GalNAc O6	Asp-125 OD1	3.02
GalNAc O7	Gly-1 NH	3.09
Gal O6	W-40	2.73
W-23	Asp-125 OD1	2.85
Ser-119 OH	Asp-125 OD2	2.65
W-23	Thr-79 O	3.15
W-40	Thr-79 NH	3.06
W-23	W-40	2.68

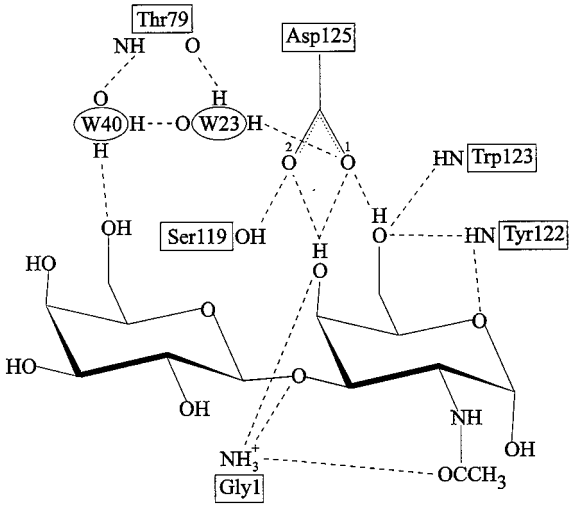

FIG. 6. **Hydrogen-bonding scheme for the MPA-disaccharide complex.** The two charged groups of Gly-1 and Asp-125 form most of the H-bonds, to the GalNAc. The remaining H-bonds are from backbone atoms plus two water molecules, one of which forms the only H-bond to the Gal.

bound by MPA in the same position and orientation as the Me α-D-Gal in the jacalin structure (15). This supports the Asp-125 carboxylate being in the charged form despite the lower pH required for the MPA crystallization, 4.5 compared with 7.3 for jacalin (15). The spectroscopic and binding properties of jacalin were interpreted as suggesting independent subsites for α and β methyl galactosides (6). However, the lack of H-bonds between the Gal and neighboring residues in the same monomer and the few van der Waals contacts suggest that there is no

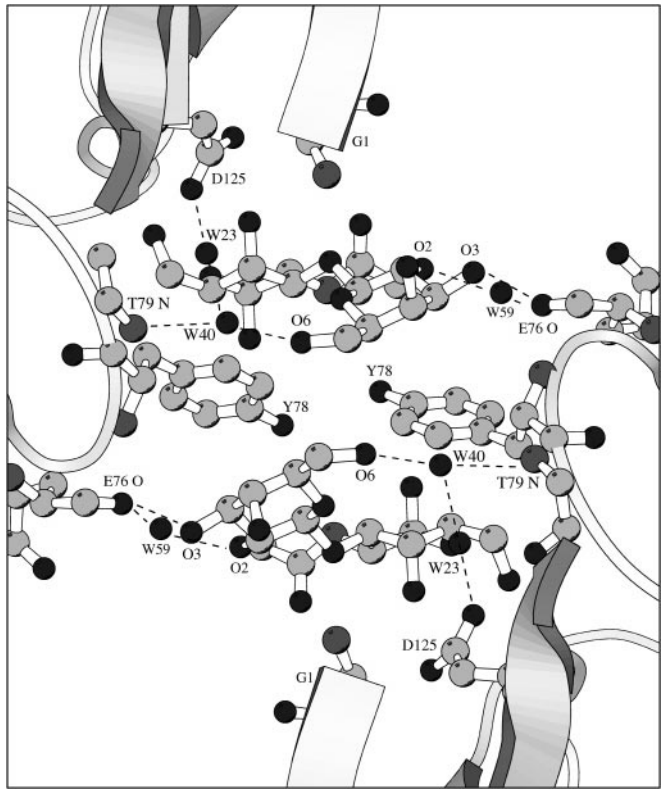

FIG. 7. **Intertetramer interactions around the Gal moieties.** The Gal O-6 binds to one tetramer through a pair of water molecules. The Gal O-2 and O-3 interact with the carbonyl oxygen of Glu-76 of the second tetramer, both by direct H-bonding and through a third water molecule. The two Tyr-78 hydroxyls form a H-bond in the center of the interfacial structure.

true subsite for the Gal moiety. It remains possible that β-galactosides could be bound in the α-Gal/GalNAc site but in an orientation that produces different spectroscopic effects. Further x-ray studies should resolve this point. The hydrogen bonding interactions of the jacalin with Me α-D-Gal and MPA with the GalNAc unit of the disaccharide are generally similar. The additional energy of binding of GalNAc compared with Gal by MPA comes in part from the H-bond from the carbonyl oxygen of the NAc to the terminal amino group (Gly-1), and this adds to the importance of this group in the overall recognition process. The methyl moiety of the acetyl group projects away from any side chains, unlike the prediction from jacalin-ligand modeling (15).

All of the ligand-contacting residues in the two proteins are completely conserved. However, the two proteins differ in two aspects of their binding. Jacalin has a three-fold higher affinity for Me α-D-Gal and shows a smaller increase in intrinsic fluorescence on binding this ligand than does MPA (7). This points to some minor conformational difference in the ligand-free structures, possibly due to residue 2 in the α-chain being Lys in jacalin and Val in MPA. These residues are in the neighborhood of Trp-123.

Comparison to Other Lectin Structures—The internal three-fold pseudo-symmetry seen in MPA also occurs in ricin (28), amaranthin (3), and snowdrop lectins (29), each of which have, however, different overall architectures to MPA. The higher internal sequence homology of the snowdrop lectin leads to a valance of 3 for its protomer, and in the ricin case the original assignment of one Gal-binding site per protomer has been challenged by recent site-directed mutagenesis experiments (30). Unlike these plant lectins, the MPA binding site is formed primarily from contiguous loops joining β-strands 1 and 2 in each Greek key subdomain; hence, it cannot have further binding sites.

The recently reported new member of the Moraceae lectin structural family, Calsepa, shows approximately 30% sequence homology to the jacalin and MPA α-chains (11). The homology is particularly strong in the second and third subdomains. More importantly, the conserved residues are mainly in the β-strands and include two of the three internal Ile residues that fill the domain interior, the third being Leu in Calsepa. Of the residues that form the binding site in MPA, Calsepa has identical residues for Asp-125 and Tyr-122, whereas Phe-47 in MPA is Ser and Trp-123 is Tyr. The N-terminal Gly can be aligned with an internal Gly in Calsepa, but there is a deletion at the equivalent position to the four-residue linker segment of jacalin (11). The conservation of the key Asp-125, which forms H-bonds with the O-4 and O-6 of the GalNAc, in a lectin with Man/maltose specificity is reminiscent of the conserved Asp/Asn feature of the binding sites of legume lectins with similar specificity differences (31). The change of Phe-47 and the extension on the N-terminal Gly will create a quite different structure in this half of the site in Calsepa, consistent with the differences at two hydroxyls of its ligands from the Gal ones of MPA. This protein also shows that the post-translational processing that appears to be essential in jacalin and MPA is not obligatory in all members of this family. Calsepa also lacks a pro-segment equivalent to the 39-residue portion preceding the β-chain in jacalin (11).

The role of the Gal interactions in the tetramer-tetramer association in the crystal packing is remarkable. It has some significant parallels in the formation of lattices by legume lectins, recently reviewed by Brewer (32), S- and C-type animal lectins (33), and the snowdrop lectin (29) when they interact with branched oligosaccharides. These lattices have been observed in both electron microscopy and x-ray crystallographic experiments with divalent, trivalent, and tetravalent lectins interacting with divalent ligands. This phenomenon is thought to be of considerable biological significance for lectin function, where cross-linking of cell-surface glycoconjugates leads to signal transduction effects. Crystals of the lattice type lectin complexes have higher solvent contents than usual and open channels, a feature also seen in the MPA case, although the solvent content is only 57%. In MPA, the placement of the sites at the corners of the tetramer is ideal for lattice formation of the kind seen with the other animal and plant lectins. However, its ability to form lattices with a small "monovalent" ligand is novel. Whether the interactions are sufficiently strong to mediate tetramer-tetramer interactions by the disaccharide in solution is presently being investigated.

In conclusion, the present MPA-disaccharide structure has shown that the recognition of the T-antigen on tumor mucins will be predominantly through the GalNAc moiety, but the geometry of the interaction may be considerably more complex if lattice formation can occur under the conditions of cell-lectin binding. The ability of MPA to differentiate lung cell tumor types (34, 35), to preferentially label human translational cell carcinomas (36) and its toxicity toward insect larvae (37) may well rest upon this unusual association mechanism.

Acknowledgments—We thank Dr. Bryan G. Williams for his continued support of this work, Dr. J.-R. Brisson for providing the minimum energy plot for the disaccharide, Rebecca To for protein purification, Dr. Stephen Evans for assistance with graphics, Dr. Focco van den Akker for initial T-antigen coordinates, Dr. A. P. Hammersley for correction of the II/CCD data, Dr. Eldon Walker for computer support, Drs. Eric Fançon and Michel Roth for beamtime at ESRF, and EMBL and ESRF for their support.

REFERENCES

1. Springer, G. F. (1995) *Crit. Rev. Oncogen.* **6,** 57–85
2. Ravishankar, R., Ravindran, M., Suguna, K., Surolia, A and Vijayan, M. (1997) *Curr. Sci.* **72,** 855–861
3. Transue, T. R., Smith, A. K., Hanqing, M., Goldstein, I. J., and Saper, M. A. (1995) *Glycoconj. J.* **12,** 425
4. Sharma, V., Vijayan, M., and Surolia, A. (1996) *J. Biol. Chem.* **271,** 21209–21213
5. Sarkar, M., Wu, A. M., and Kabat E. A. (1981) *Arch. Biochem. Biophys.* **209,** 204–218
6. Mahanta, S. K., Sastry, M. V. K., and Surolia, A. (1990) *Biochem. J.* **265,** 831–840
7. Young, N. M., Johnston, R. A. Z., Szabo, A. G., and Watson, D. C. (1989) *Arch. Biochem. Biophys.* **270,** 596–603
8. Young, N. M., Johnston, R. A. Z., and Watson, D. C. (1991) *FEBS Lett.* **282,** 382–384
9. Young, N. M., Watson, D. C., and Thibault, P. (1995) *Glycoconj. J.* **12,** 135–141
10. Yang, H., and Czapla, T. H. (1993) *J. Biol. Chem.* **268,** 5905–5910
11. Suresh, S., Rani, P. G., Pratap, J. V., Sankaranarayanan, R., Surolia, A., and Vijayan, M. (1997) *Acta Crystallogr. Sec. D* **53,** 469–471
12. Van Damme, E. J. M., Barre, A., Verhaert, P., Rougé, P., and Peumans, W. (1996) *FEBS Lett.* **397,** 352–356
13. Lee, X., Johnston, R. A. Z., Rose, D. R., and Young, N. M. (1989) *J. Mol. Biol.* **210,** 685–686
14. Banerjee, R., Dhanaraj, V., Mahanta, S. K., Surolia, A., and Vijayan, M. (1991) *J. Mol. Biol.* **221,** 773–776
15. Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia and Vijayan, M. (1996) *Nat. Struct. Biol.* **3,** 596–603
16. Hammersley, A. P. (1994) *FIT 2D Manual*, ESRF, Grenoble, France
17. Otwinowski, Z. (1993) in *Proceedings of the CCP4 Study Weekend: Data Collection and Processing* (Sawyer, L., Isaacs, N., and Bailey, S., eds) pp. 56–61, Daresbury Laboratory, Warrington, United Kingdom
18. Collaborative Computational Project No. 4 (1994) *Acta Crystallogr. Sec. D* **50,** 760–763
19. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sec. A* **47,** 110–119
20. Brünger, A. T. (1993) *X-PLOR*, version 3.1, Yale University, New Haven, CT
21. Kraulis, P. (1991) *J. Appl. Crystallogr.* **24,** 941–950
22. Merrit, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr. Sec. A* **47,** 869–873
23. Nicholls, A., and Honig, B. (1993) *GRASP: Graphical Representation and Analysis of Surface Properties*, Columbia University, New York
24. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. H. (1993) *J. Appl. Crystallogr.* **26,** 548–558
25. Bush, C. A., and Feeney, R. E. (1986) *Int. J. Pept. Protein Res.* **28,** 386–397
26. Pirie-Shepherd, S. R., Stevens, R. D., Andon, N. L., Enghild, J. J., and Pizzo, S. V. (1997) *J. Biol. Chem.* **272,** 7408–7411
27. Fersht, A. R., Shi, J.-P., Knill-Jones, J., Lowe, D. M., Wilkinson, A. J., Blow, D. M., Brick, P., Carter, C., Waye, M. M. Y., and Winter, G. (1985) *Nature* **314,** 235–238
28. Rutenbar, E., Ready, M., and Robertus, J. D. (1987) *Nature* **326,** 624–626
29. Wright, C. S., and Hester, G. (1996) *Structure* **4,** 1339–1352
30. Frankel, A., Tagge, E., Chandler, J., Burbage, C., and Willingham, M. (1996) *Protein Eng.* **9,** 371–379
31. Young, N. M., and Oomen, R. P. (1992) *J. Mol. Biol.* **228,** 924–934
32. Brewer, C. F. (1996) *Chemtracts Biochem. Mol. Biol.* **6,** 165–179
33. Gupta, D., Kaltner, H., Dong, X., Gabius, H.-J., and Brewer, C. F. (1996) *Glycobiology* **6,** 843–849
34. Welle, N. K., and Karnovsky, M. J. (1989) *Am. J. Pathol.* **134,** 277–285
35. Kresch, M. J., Lwebuga-Mukasa, J., Wilson, C. M., and Gross, I. (1991) *Lung* **169,** 139–151
36. Ward, G. K., Steward, S. S., Dotsikas, G., Price, G. B., and Mackillop, W. J. (1992) *Histochem. J.* **29,** 685–694
37. Czapla, T. H., and Lang, B. A. (1990) *J. Econ. Entomol.* **83,** 2480–2485

Structure of the Complex of *Maclura pomifera* Agglutinin and the T-antigen Disaccharide, Gal β1,3GalNAc

Xavier Lee, Andrew Thompson, Zhiming Zhang, Hoa Ton-that, John Biesterfeldt, Craig Ogata, Lulu Xu, Rosemary A. Z. Johnston and N. Martin Young

J. Biol. Chem. 1998, 273:6312-6318.
doi: 10.1074/jbc.273.11.6312
