Redefinition of the Carbohydrate Specificity of *Erythrina corallodendron* Lectin Based on Solid-Phase Binding Assays and Molecular Modeling of Native and Recombinant Forms Obtained by Site-Directed Mutagenesis†

Ernesto Moreno,‡,§ Susann Teneberg,‡ Rivka Adar,|| Nathan Sharon,|| Karl-Anders Karlsson,‡ and Jonas Ångström*,‡

Institute of Medical Biochemistry and Microbiology, Department of Medical Biochemistry, Göteborg University, Medicinaregatan 9A, S-413 90 Göteborg, Sweden, Center of Molecular Immunology, P.O. Box 16040, Havana 11600, Cuba, and Department of Membrane Research and Biophysics, The Weizmann Institute of Science, Rehovot 76100, Israel

Received September 3, 1996; Revised Manuscript Received December 5, 1996⊗

ABSTRACT: Binding of the *N*-acetyllactosamine-specific lectin from *Erythrina corallodendron* (ECorL) to four glycosphingolipids has been tested using the microtiter well assay. The role of several amino acids in the binding site region was studied by combining binding assays and molecular modeling for native and recombinant forms of the lectin. Seven single-point mutants at positions 106 (Y106A), 108 (Y108A, T), 218 (A218G), and 219 (Q219A, N or E) were investigated. A comparison with more than 30 known sequences of legume lectins showed that ECorL is unique in displaying a tyrosine residue or a structural equivalent at position 106. Analyses of the binding results obtained for mutants at positions 106 and 108 using molecular modeling point to complex conformational dependencies between these and several other residues around the binding site. Gln 219 was found to have a large conformational flexibility, which, paradoxically, favors the binding of *N*-acetyllactosamine-containing glycosphingolipids. Particularly significant is the fact that ECorL exhibits a higher affinity for Fucα2Gal*β*4GlcNAc*β*-terminated glycosphingolipids than *N*-acetyllactosamine-terminated ones, in accordance with molecular modeling revealing a perfect fit of the α2-linked fucose in a cavity extending from the Gal*β*4 binding pocket. These findings lead to a redefinition of the specificity of this lectin, where the affinity for the terminal Fucα2Gal*β*4GlcNAc*β* trisaccharide should be considered in the first place. The possible biological significance of this specificity remains to be investigated.

The name lectin is used to designate all sugar binding proteins of nonimmune origin, whether from plants, animals, or microorganisms. Among plant lectins, most of the attention has been focused on members of the *Leguminosae* family. More than 40 lectin sequences of this family are known, and the crystal structures of about ten different of these lectins have been determined, most of them in complex with carbohydrates. It has been shown that legume lectins exhibit considerable sequence (Sharon & Lis, 1990) and structural (Rini, 1995) homology, in spite of their different carbohydrate specificities, the latter being determined by relatively few residues concentrated in six peptide strands (Young & Oomen, 1992). These characteristics, together with the availability of structural data, make the family of legume lectins very suitable for studies of protein−carbohydrate interaction at the atomic level.

In terms of carbohydrate specificity, lectins are commonly classified according to the mono- or sometimes disaccharide that is the best inhibitor of erythrocyte agglutination, which is the method routinely employed for their detection and characterization. However, these proteins often demonstrate binding preferences for oligosaccharides, showing association constants that are by 2 or 3 orders of magnitude higher than those displayed for monosaccharides (Sharon & Lis, 1989). Since lectins are widely employed in glycoconjugate research and clinical laboratories, the accurate determination of their specificities is very important, especially when they are to be used as tools for characterization of binding structures. To achieve this, a battery of different oligosaccharide structures, as wide as possible, should be used. If applied in the early stages of detection, this procedure may lead to the discovery of previously nondetected lectins. Furthermore, the determination of new binding specificities for already known lectins may provide some clues toward the biological roles of these proteins, since a particular binding preference may be connected to a specific function *in vivo*.

Previous binding studies on *Erythrina corallodendron* (ECorL)[1] and *Erythrina cristagalli* (ECL) lectins revealed that both proteins recognize Fucα2Gal*β*4GlcNAc*β*-terminated glycosphingolipids as, *e.g.*, the H5 type 2 glycosphingolipid (Fucα2Gal*β*4GlcNAc*β*3Gal*β*4Glc*β*1Cer) (Teneberg *et al.*, 1994). It was also shown that ECL accommodates this terminal Fucα2 residue without any change in binding

† This work was supported by grants from The Swedish Medical Research Council (3967 and 10435) and from the Basic Research Fund, Israel Academy of Sciences and Humanities.

* Corresponding author. Fax: +46-31-413190.
‡ Göteborg University.
§ Center for Molecular Immunology.
|| The Weizmann Institute of Science.

⊗ Abstract published in *Advance ACS Abstracts,* April 1, 1997.

[1] Abbreviations: ECorL, *Erythrina corallodendron* lectin; rECorL, recombinant ECorL; ECL, *Erythrina cristagalli* lectin; PNA, peanut agglutinin; SBA, soybean agglutinin; Con A, *Canavalia ensiformis* (concanavalin A) lectin; LOL-I, *Lathyrus ochrus* isolectin I; GS4, *Griffonia simplicifolia* isolectin IV; lactosylceramide, Gal*β*4Glc*β*1Cer; *N*-acetyllactosamine, Gal*β*4GlcNAc; fucosyl-*N*-acetyllactosamine, Fucα2Gal*β*4GlcNAc; neolactotetrasylceramide, Gal*β*4GlcNAc*β*3-Gal*β*4Glc*β*1Cer; H5 type 2 glycosphingolipid, Fucα2Gal*β*4Glc-NAc*β*3Gal*β*4Glc*β*1Cer; MD, molecular dynamics; RT, room temperature. All monosaccharides are of the D configuration except L-fucose.

S0006-2960(96)02231-3 CCC: $14.00

affinity as compared with neolactotetraosylceramide (Galβ4GlcNAcβ3Galβ4Glcβ1Cer). In the present work, we report binding and molecular modeling studies performed on the native lectin, whose crystal structure in complex with lactose has been determined at a resolution of 2.0 Å (Shaanan *et al*., 1991), and recombinant forms of ECorL, leading to a redefinition of ECorL as a Fucα2Galβ4GlcNAcβ-specific lectin.

EXPERIMENTAL PROCEDURES

Lectin Mutants. ECorL was purchased from Sigma. Recombinant ECorL and the seven mutant lectins (Q219A, Q219E, Q219N, A218G, Y106G, Y108A, and Y108T) were generated and isolated as described (Arango *et al*., 1992, Adar & Sharon, 1996); in all cases the recombinant DNAs were fully sequenced to confirm the planned mutations and to ascertain that no spurious mutations had occurred (Adar & Sharon, 1996). It may be noted that rECorL differs from ECorL at seven positions (Arango *et al*., 1990), all of which are far from the saccharide binding site except Gln 134, which replaces the proline found in the native protein. For the binding assays 100 μg of EcorL, rECorL, and the respective mutants were labeled with ^{125}I by the Iodogen method (Aggarwal *et al*., 1985), yielding in average 2×10^3 cpm/μg.

Glycosphingolipid Binding Assays. The glycosphingolipids used in the binding assays were isolated as described (Karlsson, 1987), and the identity of the purified compounds was established by mass spectrometry, proton NMR spectroscopy and degradation studies, as outlined in Teneberg *et al*. (1994). Lactosylceramide (Galβ4Glcβ1Cer) was from dog intestine (Hansson *et al*., 1983) while neolactotetraosylceramide (Galβ4GlcNAcβ3Galβ4Glcβ1Cer) and the H5 type 2 glycosphingolipid (Fucα2Galβ4GlcNAcβ3Galβ4Glcβ1Cer) were from human erythrocytes (Laine *et al*., 1974). De-*N*-acylated neolactotetraosylceramide (Galβ4GlcNH$_2\beta$3Galβ4Glcβ1Cer) was obtained by treatment with anhydrous hydrazine (Keith *et al*., 1978).

Binding of radiolabeled lectins to glycosphingolipids adsorbed in microtiter wells was done as described previously (Teneberg *et al*., 1994). In brief, serial dilutions (each in triplicate) of pure glycosphingolipids in methanol were applied in microtiter wells (Cooks M24, Nutacon, Holland) and kept at room temperature (RT) until the solvent had evaporated. To diminish unspecific protein binding, the wells were coated with 200 μL of phosphate-buffered saline (PBS), pH 7.3, containing 2% (w/v) bovine serum albumin and 0.1% NaN_3 (solvent I) for 2 h at RT. After being washed once with solvent I, radiolabeled lectins were added to the wells (50 μL/well; diluted to approximately 2×10^3 cpm/μL in solvent I), followed by incubation for 4 h at RT. The wells were then washed six times with solvent I, and after being dried, the radioactivity was counted in a γ counter.

Molecular Modeling. Molecular modeling was conducted on a Silicon Graphics Indigo^2 Extreme workstation using Quanta 4.1/CHARMm 23 software (Molecular Simulations Inc., 1995).

Protein models were constructed on the basis of the crystal structure of the complex between ECorL and lactose (Shaanan *et al*., 1991), available from the Brookhaven Protein Data Bank (Bernstein *et al*., 1977), entry 1LTE. Solvent molecules were removed, except those that were completely surrounded by the protein environment. The *N*-linked heptasaccharide of the lectin (at Asn 17) was also removed since it is far from the binding site. Subsequently, both polar and nonpolar hydrogens were added to the structure. To refine the stereochemistry and molecular contacts, a constrained energy minimization of the protein (without the ligand) was carried out. Harmonic constraints of 20 kcal/(mol Å^2) were applied to all α carbons, the two metal ions present in the lectin (Ca^{2+} and Mn^{2+}), and the oxygen atoms of the four internal water molecules which interact with them, during 100 steps of a steepest descent geometry relaxation. Afterward, harmonic constraints were reduced to 5 kcal/(mol Å^2) to perform 300 steps of conjugate gradient minimization. A distance-dependent dielectric constant ($\epsilon = 6r$) and a 15 Å cutoff distance were used for nonbonded interactions. At the end of this procedure the average force gradient was 0.15 kcal/(mol Å), and the rms deviation from the crystal structure was 0.1 Å, if calculated for main chain atoms, and 0.41 Å when calculated for all atoms in the protein. The Ca^{2+} and Mn^{2+} ions were displaced from their crystal positions about 0.05 and 0.15 Å, respectively. The obtained structure was used as starting geometry for construction of mutants, as well as for docking procedures and dynamics simulations.

Mutants were constructed using the Protein Design facilities within Quanta. The program places each new side chain in a conformation as similar as possible to the old one. After such automatic replacements the side chain conformations were adjusted, if necessary, taking into account rotamer populations and steric contacts with the rest of the protein. Finally, local energy minimization was applied within a distance of 7 Å from the mutated residue, keeping the rest of the protein fixed. The resulting conformations of neighboring side chains and backbone were very similar to the crystal structure.

Docking of Ligands. Models of the terminal disaccharide of neolactotetraosylceramide (Galβ4GlcNAcβ3) and the terminal trisaccharide of the H5 type 2 glycosphingolipid (Fucα2Galβ4GlcNAcβ3) were constructed for docking purposes. Starting values for the dihedral angles over the glycosidic bonds of the Galβ4GlcNAc and Fucα2Gal disaccharides were taken from the literature (Imberty *et al*., 1990, 1995). First, the lactose moiety present in the crystal complex was replaced by the respective sugar fragment by superimposing the galactose rings. Then, possible combinations of glycosidic angle values and amino acid side chain orientations were examined to find the optimal docking conformation in each case. Hydroxyl groups in both the protein and the ligand were oriented so as to favor hydrogen bond interactions. At the end an energy minimization was carried out for each complex.

Molecular Dynamics (MD) Simulations. MD simulations were run in vacuum and/or in a water environment. Vacuum simulations were carried out mainly to study local conformational properties of the proteins in the absence of ligand. A distance-dependent dielectric constant ($\epsilon = 6r$) and a constant one ($\epsilon = 1$) were used for vacuum and water simulations, respectively. The SHAKE algorithm as implemented in CHARMm was used to constrain bonds containing hydrogen atoms, allowing a time step of 0.001 ps. Simulations were started by heating the system from 0 K up to the final temperature (300 K unless otherwise specified) using a 5 K increment every 100 time steps, followed by an equilibration period of 10 ps.

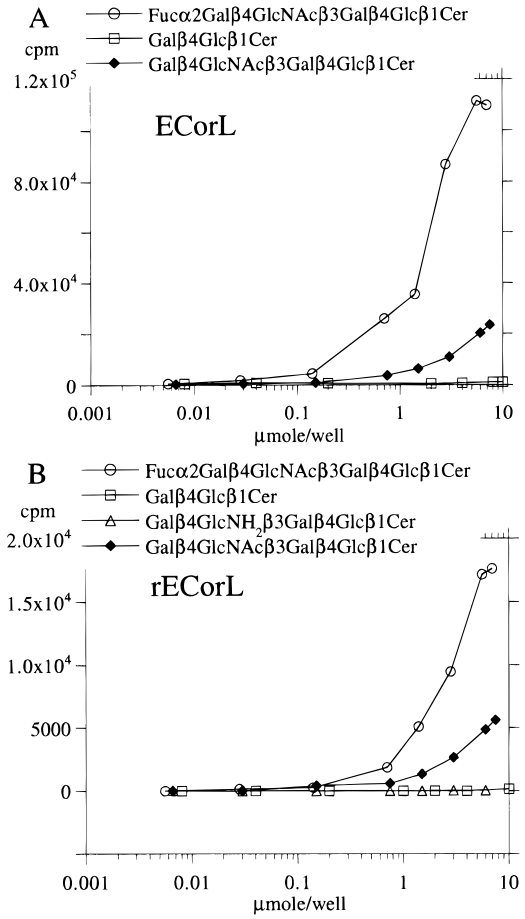

FIGURE 1: Results from binding of ^{125}I-labeled lectin from *E. corallodendron* (ECorL, panel A) and recombinant *E. corallodendron* lectin (rECorL, panel B) to serial dilutions of glycosphingolipids adsorbed in microtiter wells. The data represent mean values of triplicate determinations.

Sequence Analysis. Sequences of legume lectins were extracted from the Swiss-Prot database and subsequently analyzed using different programs implemented within the Wisconsin Sequence Analysis Package (Genetics Computer Group, Inc., 1994).

RESULTS AND DISCUSSION

Binding of ECorL and rECorL to Glycosphingolipids Adsorbed in Microtiter Wells. In a previous investigation of the binding specificities of ECorL and ECL it was demonstrated that on thin-layer plates both recognize Galβ4GlcNAcβ-terminated glycosphingolipids as, *e.g.*, neolactotetraosylceramide (Teneberg *et al.*, 1994). They also interacted with glycosphingolipids having a terminal blood group H type 2 determinant (Fucα2Galβ4GlcNAcβ) and thus tolerated a substitution with a Fucα at the 2-position of the Galβ4. Using the microtiter well binding assay it was further demonstrated that ECL bound with similar affinity to neolactotetraosylceramide and the H5 type 2 glycosphingolipid.

However, binding of ^{125}I-labeled ECorL to glycosphingolipids adsorbed in microtiter wells (Figure 1A) resulted in a preferential binding to the H5 type 2 glycosphingolipid. The lectin also bound to neolactotetraosylceramide at a level estimated to be at least 5 times lower than that of the H5 type 2 glycosphingolipid, while lactosylceramide was nonbinding. The same binding pattern was obtained for rECorL (Figure 1B). This assay also showed that the interaction of rECorL with neolactotetraosylceramide was abolished by conversion of the acetamido group of GlcNAc to an amine (Galβ4GlcNH$_2\beta$3Galβ4Glcβ1Cer). These results will be discussed below in conjunction with molecular modeling analyses.

It is to be noted that free lactose inhibits hemagglutination of either protein better than galactose by a factor of 3 (Adar & Sharon, 1996) and that Fucα2Galβ4Glc inhibits hemagglutination of ECorL 15 times better than galactose (Adar and Sharon, unpublished), pointing to a contribution to binding by fucose of approximately a factor of 5 as also found above. However, the lack of lactosylceramide binding in the microtiter well assay indicates that the sensitivity of this assay is lower than in the hemagglutination assay, a finding which mainly may be ascribed to the fact that a distribution of conformations around the GlcβCer linkage is at hand, resulting in an unfavorable binding epitope presentation for several conformers of this glycosphingolipid. Furthermore, previous experiments using radiolabeled glycosphingolipids demonstrate that no selective losses occur under these assay conditions (Strömberg, 1988).

Docking of Sugar Fragments to ECorL. The docking of Galβ4GlcNAcβ and Fucα2Galβ4GlcNAcβ was carried out as described above. For both ligands, the binding site of ECorL accepts conformations that are within the global energy minimum region, while conformations belonging to other local minima are sterically forbidden. Energy minimization of the complexes did not produce important deviations from the starting geometries. The resulting interactions between the galactose residue and the protein were in both cases essentially the same as those found in the crystal structure of the ECorL−lactose complex (Shaanan *et al.*, 1991), as shown in Figure 2.

The fucose residue of Fucα2Galβ4GlcNAcβ fits very well into a cavity which serves as an extension of the Galβ4 binding pocket. The two cavities are connected through a narrow opening between the Asn 133 and Ala 218 side chains. Figure 2 illustrates most of the interactions of fucose with the protein. The $-NH_2$ group of Asn 133, which in the ECorL−lactose crystal structure was found to participate in a hydrogen bond to the 3-OH of Galβ4, also takes part in hydrogen bond interactions with the 2-OH of fucose and the glycosidic oxygen of the Fucα2Gal linkage. On the other hand, the 2-OH may also form a hydrogen bond with the backbone NH of Gly 107. A fourth hydrogen bond is found between the 3-OH and the hydroxyl group of Tyr 108. Furthermore, residues Pro 134 and Trp 135 make hydrophobic contacts with the 6-CH_3 of fucose, as well as with its 2- and 3-ring hydrogens, and together with Tyr 108 close the binding site from the fucose side.

In rECorL a glutamine is found at position 134 instead of proline (Arango *et al.*, 1990). This change does not cause any significant alterations of either backbone or side chain positions of the 132−137 loop in the absence of a ligand. However, the side chain of Gln 134 is not able to provide interactions with the fucose of Fucα2Galβ4GlcNAcβ, corresponding to what is observed for the proline in the native protein, but this does not appear to have any significant effect on the affinity for the H5 type 2 glycosphingolipid as judged by the binding curves in Figure 1. Consequently, this modification was ignored in subsequent modeling.

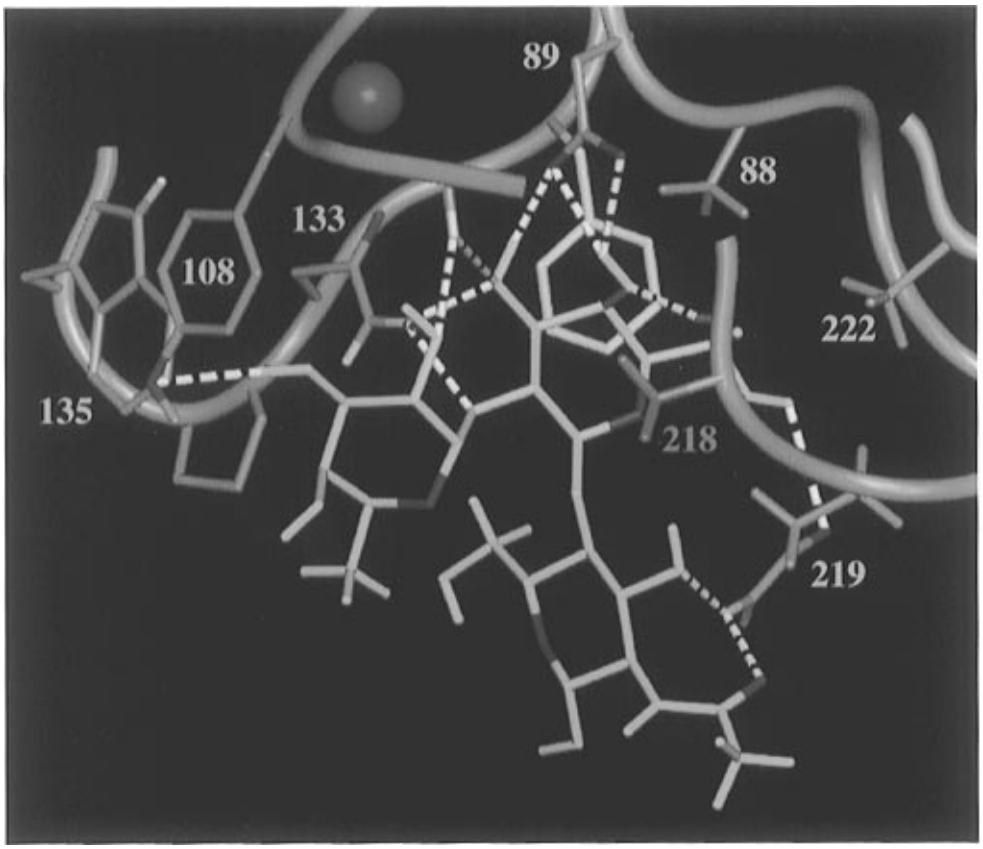

FIGURE 2: Model of the complex between ECorL and the terminal trisaccharide of the H5 type 2 glycosphingolipid (Fucα2Galβ4GlcNAcβ3), showing the hydrogen bond interactions between the sugar residues and the protein (see the text). The trisaccharide is in yellow, with ring oxygens marked in red. Most of the amino acid side chains which are relevant for the binding are shown in blue, with polar hydrogens marked in light gray and oxygens in red. The aromatic ring of Phe 131, on which galactose is stacked, is in light blue; Ala 218 in front of galactose is in green. The calcium ion is shown in magenta.

Gln 219 and the Role of the Acetamido Group of GlcNAc. It has been shown previously that *N*-acetyllactosamine inhibits the agglutination of human erythrocytes, induced by ECorL, almost 7 times better than lactose (Adar & Sharon, 1996). This difference in binding strength between the two disaccharides was ascribed to possible interactions between Gln 219 and the acetamido group of *N*-acetyllactosamine. Simple inspection and torsion angle manipulations of the model of the complex between ECorL and *N*-acetyllactosamine indicate that the acetamido group of GlcNAc should interact mostly with the side chain of Gln 219. However, the conformation adopted by the Gln 219 side chain in the crystal structure does not favor interactions between these groups that would explain the higher binding affinity of *N*-acetyllactosamine.

In order to study the possible conformations of this particular residue, a 1024 ps MD was run in vacuum for a protein region around Gln 219 in the absence of any ligand. Only residues having atoms within a radius of 6 Å from Gln 219 were allowed to move. The conformational behavior displayed by this side chain during the dynamics run is illustrated in Figure 3. Figure 3A corresponds to a scatter map of the two principal side chain torsion angles: χ_1 ($C-C_\alpha-C_\beta-C_\gamma$) and χ_2 ($C_\alpha-C_\beta-C_\gamma-C_\delta$). According to this map, there are five accessible conformational regions (from I to V). The third side chain angle χ_3 ($C_\beta-C_\gamma-C_\delta-N_\epsilon$) varied along the whole 360° range without showing any significant correlation neither with χ_1 nor with χ_2. Figure 3B shows three conformations representing the three most populated regions, which have a χ_2 angle close to 180° in common.

Region I ($\chi_1 \approx -60°$) contains the conformation observed in the crystal structure of the ECorL−lactose complex. As pointed out above, hydrogen bond interactions between the acetamido group and glutamine are not favored in this case, since the glutamine nitrogen (N_ϵ) would have to adopt an eclipsed position with respect to the β carbon (C_β) (only very few dynamics frames corresponded to this geometry, and all of them had relatively high potential energies).

Region II ($\chi_1 \approx 180°$) was the most populated one and serves as a kind of connection center for interconversions between regions. Figure 3B shows a conformation belonging to region II. Several favorable interactions with the GlcNAc residue are possible in this case, as shown in Figure 2. The glutamine $-NH_2$ can simultaneously form hydrogen bonds with the GlcNAc 3-OH oxygen and the carbonyl oxygen of the acetamido group. At the same time, the amide oxygen of Gln 219 may form a hydrogen bond with the 6-OH of Galβ4. Region III ($\chi_1 \approx 60°$) includes conformations that can display the interactions just described, except the one with the 6-OH of Galβ4. Conformations from regions IV and V, which are less populated, are not adequate for binding due to severe steric bumps with the Galβ4 residue.

The backbone angles around Gln 219 also varied during the dynamics run. These variations correlated with the conformational changes of the glutamine side chain. In the crystal structure the C_γ carbon of Gln 219 participates in a hydrophobic patch together with Ala 222 and the 6-CH_2 of

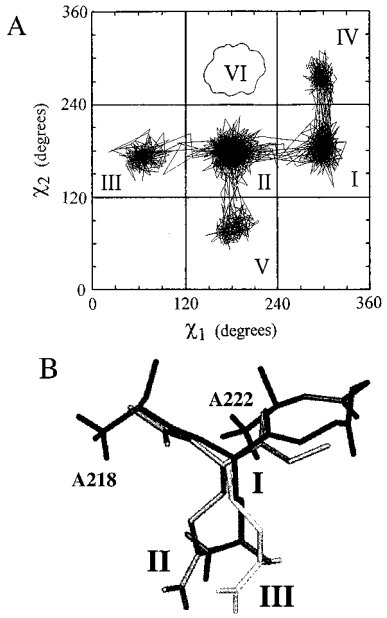

FIGURE 3: Scatter plot showing the behavior of the two principal side chain torsion angles in Gln 219, χ_1 (C−C_α−C_β−C_γ) and χ_2 (C_α−C_β−C_γ−C_δ), during a 1024 ps dynamics run in the absence of any ligand (A). Trajectory points are connected with thin lines. Five conformational regions (I−V) were obtained in this run, whereas an additional region (VI) was obtained in another simulation including solvent molecules and a ligand. Regions I, II, III, and VI can be populated if a galactose residue occupies the binding pocket. In (B) three Gln 219 side chain conformations, representative of regions I, II, and III, are shown.

Galβ4, but in conformations from regions II and III the backbone angles of Gln 219 are changed so that the C_β carbon takes the place previously occupied by C_γ (Figure 3B). The GlcNAc residue also contributes to this hydrophobic patch with the methyl group of its acetamido moiety (see Figure 2).

Another region of conformers ($\chi_1 \approx 180°$, $\chi_2 \approx -60°$) was obtained in a subsequent dynamics run that included the ligand *N*-acetyllactosamine as well as solvent molecules. The location of this region in the scatter map is also shown in Figure 3A (region VI). The interactions between the glutamine side chain and the ligand that are possible in this case are essentially the same as those described for region III.

Thus, the Gln 219 side chain shows a large flexibility favoring interactions with *N*-acetyllactosamine for most of the conformations. The most important of these interactions are hydrogen bonds between the $-NH_2$ of glutamine and the carbonyl oxygen of the acetamido group in GlcNAc and/or the 3-OH of the same residue. This may explain the lack of binding obtained for Galβ4Glc$NH_2\beta$3Galβ4Glcβ1Cer: the amine substituting the acetamido group would be repelled by the glutamine $-NH_2$.

Q219E, Q219N, and Q219A Mutants. The three different substitutions of Gln 219 used here (Glu, Asn, or Ala) all severely affected the binding to both neolactotetraosylceramide and the H5 type 2 glycosphingolipid (Figure 4), thus demonstrating the importance of this residue for binding. However, for the H5 type 2 glycosphingolipid some residual binding in the case of the Q219A and Q219N mutants is discernible, but quantitation is not possible (Figure 4B). These results are consistent with those from hemagglutination experiments showing that the minimal hemagglutinating

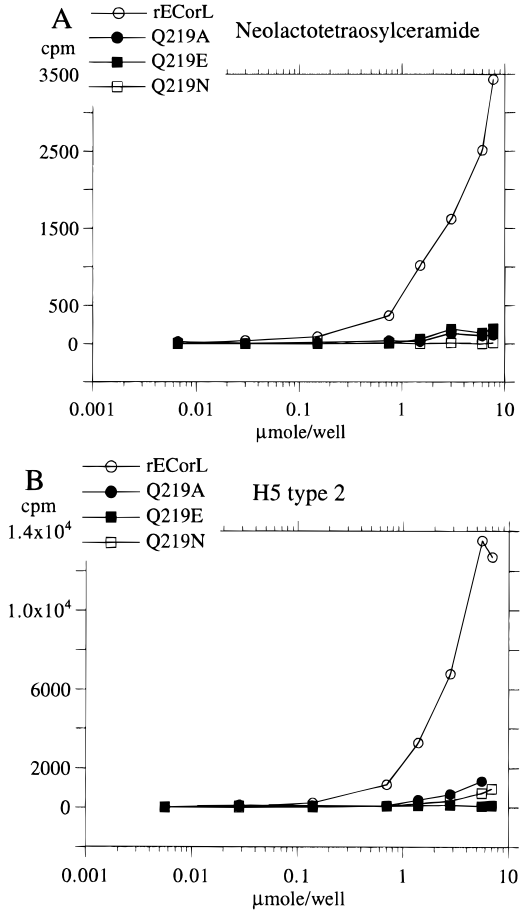

FIGURE 4: Results from binding of ^{125}I-labeled rECorL, and mutants thereof in which Gln 219 had been substituted with Ala (Q219A), Glu (Q219E), and Asn (Q219N), respectively, to serial dilutions of neolactotetraosylceramide (A) and the H5 type 2 glycosphingolipid (B) adsorbed in microtiter wells. The data represent mean values of triplicate determinations.

lectin concentration increases from 3 μg mL^{-1} for rECorL to 24 and 60 μg mL^{-1} for the Q219A and Q219N mutants, respectively, whereas the Q219E mutant was found to be inactive also in this assay (Adar & Sharon, 1996). In the case of the latter mutant electrostatic repulsion between the negatively charged carboxyl group of the glutamic acid side chain and the acetamido carbonyl oxygen of GlcNAcβ3 most likely accounts for the inactivity. However, considering the strategic position of this residue, located just at the entrance of the binding pocket, it is also conceivable that the unfavorable electrostatic influence of Glu 219 could be important already at an early stage when the ligand is approaching the protein. In this respect, it is worthwhile noting that glutamic acid at positions equivalent to 219 in ECorL was not found within a group of more than 30 legume lectin sequences (see Table 1).

The situation is different for the Q219N and Q219A mutants. Neither asparagine nor alanine can interact with the acetamido group of GlcNAcβ3 due to the shorter length of their side chains. Asparagine may, however, form a hydrogen bond with the 6-OH of Galβ4 whereas alanine may contribute to the above-mentioned hydrophobic patch involving Ala 222 and the 6-CH_2 of Galβ4, thus explaining the low but still significant affinities of these two mutants for the H5 type 2 glycosphingolipid.

Binding to Other Mutants. (A) A218G. The large reduction in binding affinity for neolactotetraosylceramide upon

Table 1: Alignment of Legume Lectin Sequences for Segments Containing Residues Subjected to Site-Directed Mutagenesis in ECorL[a]

		106				218			
		|	|			|	|		
*	ECorL	AQG	YGY	LGI	....	GATG	AQRDA		AETH
*	SBA	QTH	AGY	LGL	....	AATG	LDI-P		GESH
	DBL	RRN	GGY	LGV	....	ATTG	LSEGY		IETH
	DB58	KSN	SGF	LGV	....	ATTG	FFEGY		TETH
	PHA-L	KDK	GGF	LGL	....	ATTG	INKGN		VETN
	PHA-E	KDK	GGL	LGL	....	ATTG	ITKGN		VETN
	LBL	KKK	GRL	LGL	....	ATSG	A----		YETH
	DLL	SYH	GGF	LGL	....	ASTG	Q---N		IERN
*	LOL-I	QTG	GGY	LGV	....	ATFG	A---E		FAAH
	LSL	RGD	GGL	LGV	....	ASTA	TY---		YSAH
*	LCL	QTG	GGY	LGV	....	ATTG	A---E		FAAQ
*	PSA	QTG	GGY	LGV	....	ATTG	A---E		YAAH
	VFL	QTG	GGY	LGV	....	ATTG	A---E		YATH
*	Con A	GST	GRL	LGL	....	ASTG	L---Y		KETN
	DGL	GSG	GRL	LGL	....	ATTG	L---Y		KETN
	BMA	GSG	GRL	LGL	....	ATTG	Q---Y		TQTN
*	PNA	SIG	GGT	LGV	....	ASGS	L--GG		RQIH
	SL	KSG	GGY	LGI	....	AATG	---DL		VEQH
	UEA-I	RRA	GGY	FGL	....	GGTY	I--GR		QATH
	UEA-II	GSS	AGM	FGL	....	GGVG	N--AA		KFDH
	CSII	QSA	GGY	LGL	....	ATTG	QTDNY		IETH
	LAA-I	GSS	AGM	FGL	....	AGVG	N--AA		KFNH
	LTA	GST	GGF	LGI	....	ATTG	N--PE		REKH
	MTL-I	IGR	AGF	LGV	....	AATG	A---E		FAEH
	MTL-II	IHH	GGY	LGV	....	SSTG	A---E		YSAH
	BPA	KDY	GGC	LGL	....	GGTG	----F		NETQ
*	GS4	KDY	GGF	LGL	....	AGVG	----Y		DEVT

[a] Comparison of amino acid sequences of legume lectins for segments 103−111 and 214−226 (ECorL numbering), which contain the positions that were subjected to site-directed mutagenesis in ECorL (106, 108, 218, and 219, marked with |). More than 30 sequences were analyzed, 27 of which are included in the table (repeated sequences were omitted). Gaps (−) were introduced for optimal alignment. The two segments are separated by (....). An * indicates the availability of crystal structure(s). Sequences are grouped by suborders and tribes. Lectin abbreviations other than those given previously are as follows: DBL and DB58, *Dolichos biflorus*; PHA-L, *Phaseolus vulgaris* leucoagglutinin; PHA-E, *P. vulgaris* erythroagglutinin; LBL, *Phaseolus lunatus*; DLL, *Dolichus lab lab*; LSL, *Lathyrus sphaericus*; LCL, *Lens culinaris* (lentil); PSA, *Pisum sativum* (pea); VFL, *Vicia faba*; Con A, *Canavalia ensiformis*; DGL, *Dioclea grandiflora*; BMA, *Bowringia mildibraedii*; PNA, *Arachis hypogaea* (peanut); SL, *Onobrychis viciifolia*; UEA-I and UEA-II, *Ulex europeus* isolectins I and II, respectively; CSII, *Cytisus scoparius* lectin II; LAA-I, *Laburnum alpinum* lectin I; LTA, *Lotus tetragonolobus*; MTL-I and MTL-II, *Medicago truncatula* isolectins I and II, respectively; BPA, *Bauhinia purpurea*.

substitution of Ala 218 by glycine (Figure 5A) shows the importance of this residue in keeping the galactose ring tightly within the binding pocket. A similar role is played by leucine in the soybean lectin, as shown in the crystal structure of its complex with an analog of the blood group I carbohydrate antigen (Dessen *et al*., 1995). In this lectin, the loop containing Leu is one amino acid shorter than the corresponding loop in ECorL, leading to a different backbone geometry, but the larger side chain of Leu compensates for this difference.

As was pointed out above, the binding site of ECorL has a narrow side opening between the Asn 133 and Ala 218 side chains, which allows some substitutions to be made at C-2 of Galβ4 as, *e.g*., by a dansyl moiety (Arango *et al*.,

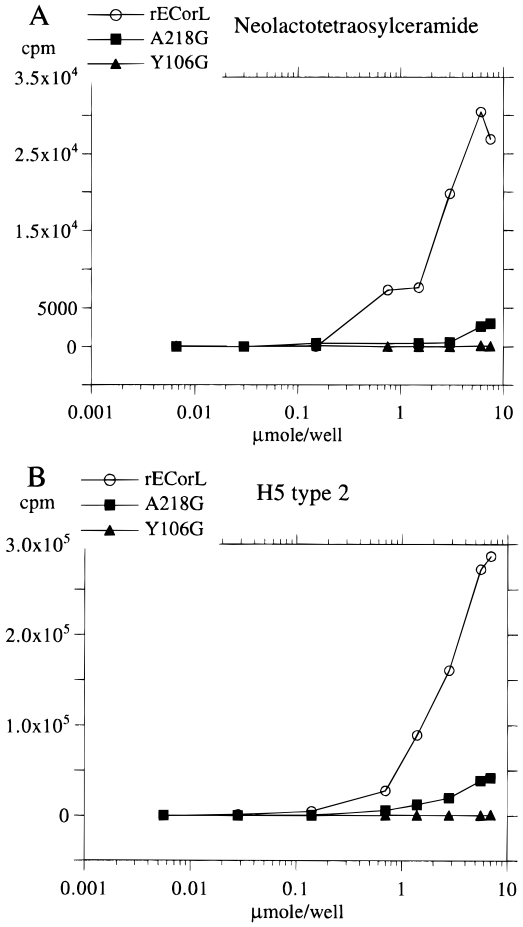

FIGURE 5: Results from binding of ^{125}I-labeled rECorL, and mutants thereof in which Ala 218 was substituted with Gly (A218G) and Tyr 106 was substituted with Gly (Y106G), to serial dilutions of neolactotetraosylceramide (A) and the H5 type 2 glycosphingolipid (B) adsorbed in microtiter wells. The data represent mean values of triplicate determinations.

1993). Nevertheless, this aperture is narrow enough to restrict displacement of the ligand in this direction. When alanine is replaced by glycine, the opening becomes wider and the galactose residue will not be bound as tightly as in the native protein. The binding to the terminal trisaccharide of the H5 type 2 glycosphingolipid is affected almost to the same degree by this mutation, as shown in Figure 5B, indicating that addition of fucose stabilizes this complex only to a minor degree.

(B) Y106G. The absence of binding activity exhibited by this mutant (Figure 5) was quite surprising at first since most legume lectins with different binding specificities have glycine at this position and several of them present the same structural motif as this mutant, characterized by the presence of three consecutive glycines in a hairpin loop that points to the binding pocket and interacts from one side with the metal ions (see Table 1). The crystal structures of three of these lectins, lentil (Loris *et al*., 1993), pea (Einspahr *et al*., 1986), and LOL-I (Bourne *et al*., 1990), display the same loop geometry as in ECorL, strongly suggesting that the mutation of Tyr by Gly should not affect the conformation of this particular loop. In agreement with this, several MD simulations carried out for the mutated loop did not produce any important conformational changes. On the other hand, Tyr 106 in ECorL participates only to a minor degree in the binding interactions with either neolactotetraosylceramide or the H5 type 2 glycosphingolipid, so no remarkable variations

in the binding properties should be expected in this case, unless other major changes occur in the structure.

Other possible conformational changes in the vicinity of residue 106 were subsequently investigated. The closest amino acid is Ala 218, which makes hydrophobic contacts with the $C_\beta H_2$ group and part of the phenyl ring of Tyr 106 in the native protein. In ECorL, Ala 218 is contained within a loop which is two or three residues longer than the corresponding loops in the three lectin structures mentioned above. This increase in length could lead to a larger flexibility and therefore to different conformational arrangements of the loop.

A 400 ps MD simulation was therefore run at 350 K in which a long segment containing the loop of interest was allowed to move (from amino acid 214 to 224) as well as neighboring residues. The loop containing Gly 106 was kept fixed to avoid unrealistic distortions which could arise because of the higher temperature. The starting conformation was very close to the one present in the crystal structure of ECorL. The loop containing Ala 218 (residues 218−222) experienced large movements. After the first 130 ps a major conformational change occurred in which Ala 218 turned into the binding pocket, facing the ring of Phe 131. This overall conformation remained without major changes for the rest of the dynamics run (270 ps). The transition was accomplished by large movements in which the side chain of Ala 218 and some backbone atoms passed through positions that were occupied by the $-CH_2$ group and part of the phenyl ring of Tyr 106 in the native protein and which normally would prevent such a transition to take place. Similar simulations were carried out for native ECorL, using even higher temperatures, but no conformational changes of this kind were observed. These modeling studies thus suggest that Tyr 106 plays an important role in supporting the conformation of the neighboring 218−222 loop containing residues critical for binding. Adar and Sharon (1996) obtained a weak hemagglutination activity for the Y106A mutant, suggesting that the methyl group of alanine to some degree is able to restrict the conformational transition described above.

Comparison with more than 30 other known legume lectin sequences reveals that Tyr 106 in ECorL is unique (Table 1). In the rest of the sequences, only small amino acids are present at structurally equivalent positions, glycine and alanine being the most common ones, whereas serine is found in two lectins. It should be noticed that, for most of the sequences, the segments equivalent to the 218−222 loop in ECorL are shorter by two or more amino acids, from which less conformational flexibility should be expected. On the other hand, for several lectins of the *Phaseoleae* tribe (first group of sequences in Table 1), in which the corresponding loops are of the same length as in ECorL, and where the positions equivalent to 218 are occupied by large hydrophobic amino acids, the conformational change described above should not be possible. Finally, inspection and simple modeling manipulation of several crystal structures of lectins displaying short loops, corresponding to 218−222 in ECorL, such as LOL-I, lentil lectin, Con A (Weisgerber & Helliwell, 1993), and GS4 (Delbaere *et al*., 1993), indicate that the presence of tyrosine or other large side chains at positions equivalent to 106 in ECorL might result in complete loss of binding activity via blocking of the binding site.

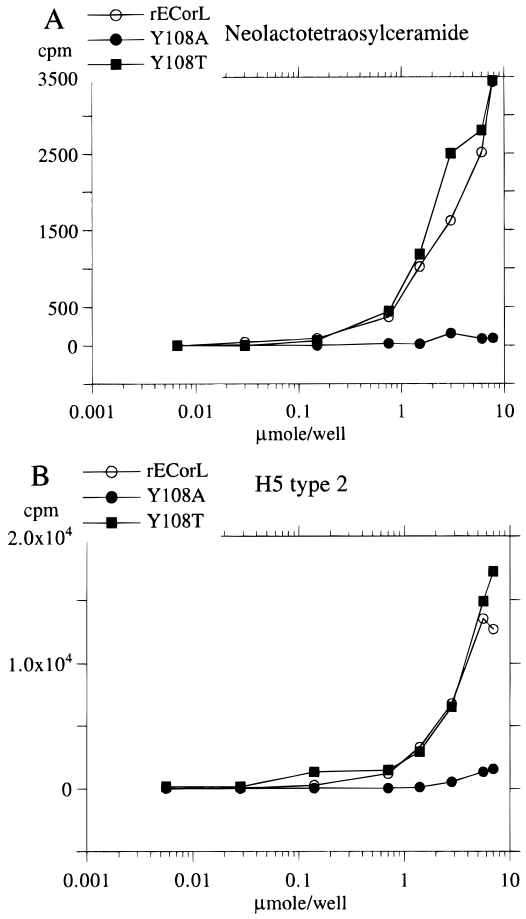

FIGURE 6: Results from binding of ^{125}I-labeled rECorL, and mutants thereof in which Tyr-108 was substituted with Ala (Y108A) and Thr (Y108T), to serial dilutions of neolactotetraosylceramide (A) and the H5 type 2 glycosphingolipid (B) adsorbed in microtiter wells. The data represent mean values of triplicate determinations.

(C) Y108A. The replacement of Tyr 108 by Ala abolished binding of both neolactotetraosylceramide and the H5 type 2 glycosphingolipid in the microtiter well assay (Figure 6). However, hemagglutination inhibition experiments reveal a drop in relative affinity of Y108A for Fucα2Galβ4Glc only by a factor of 2 as compared to ECorL and a minimal hemagglutinating concentration increased by a factor of 4 (Adar and Sharon, unpublished), again reflecting the different sensitivities of the two assay systems. As shown above, Tyr 108 contributes to a hydrogen bond when binding to the terminal trisaccharide of the H5 type 2 glycosphingolipid but is distant from the Galβ4 binding zone. Thus, major conformational changes must occur in order to account for the loss of binding activity for the two glycosphingolipids and the lowered relative affinity of Fucα2Galβ4Glc.

Comparison with sequences of other legume lectins shows that positions equivalent to 108 in ECorL always exhibit a large hydrophobic residue (usually Tyr or Phe), the only exception being PNA, which has Thr instead (see Table 1). The next two residues in the sequence are also very conserved: a large hydrophobic amino acid (Leu in most cases, with the exception of three lectins having Phe), followed always by Gly. These three consecutive residues appear to be important in keeping a backbone turn present in all solved crystal structures. Substitution of Tyr 108 by Ala might thus affect the backbone conformation of this region, which in turn could affect the binding pocket.

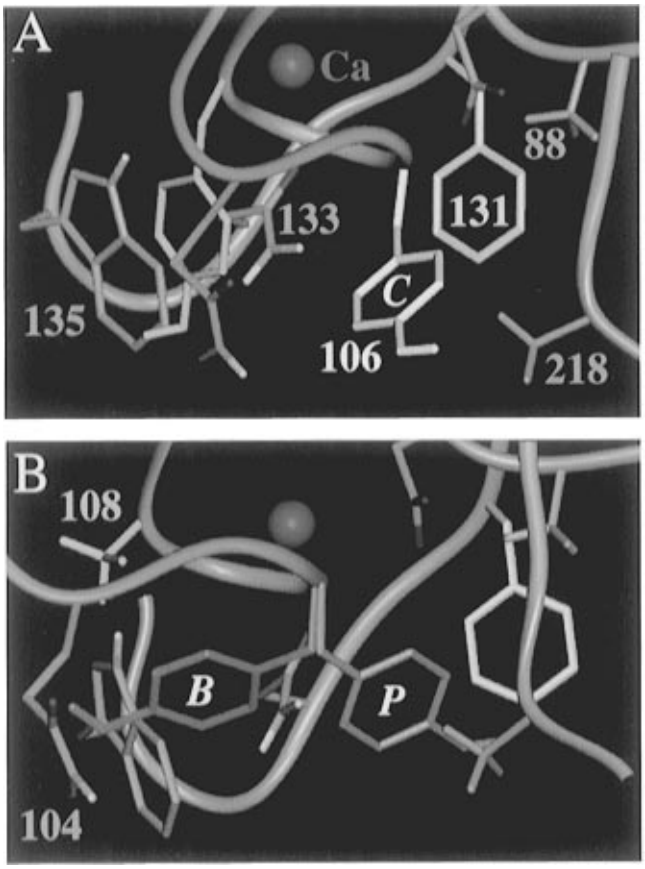

FIGURE 7: Conformational possibilities for Tyr 106 in ECorL and the Y108A mutant. Conformation *C* is observed in the *C*rystal structure of ECorL (A). Tyr 108 is shown in yellow. The Y108A mutant allows conformation *B* (in green, *B*elow Ala 108) from which transitions to conformation *P* (in pink) that blocks the binding *P*ocket easily may occur (B). In both panels Phe 131 is in light blue; polar hydrogens on side chains other than Tyr 106 are marked in light gray and oxygens in red. Both panels show the same amino acid residues, their labels being complementary.

Molecular modeling suggests another possible effect involving the Tyr 106 side chain. There are three staggered positions which a tyrosine ring in principle may adopt due to rotation around the $C_\alpha-C_\beta$ bond as shown in Figure 7. Conformation *C* in Figure 7A is the one observed in the *C*rystal structure. Conformation *P*, shown in pink in Figure 7B, places this side chain within the binding *P*ocket in a very favorable position between residues Phe 131 and Ala 218. A third conformation, which is forbidden in ECorL, becomes possible for the Y108A mutant since the small Ala residue accepts the tyrosine ring *B*elow it (conformation *B* in Figure 7B), making close contacts with the methyl hydrogens. Interconversions between conformations *B* and *C* in Y108A are very improbable due to steric hindrance from Gln 104. On the other hand, once Tyr 106 has been placed in conformation *C*, a direct transition to *P* is strictly forbidden in both the native protein and the mutant due to severe steric hindrance from Ala 218. Energy calculations were carried out for these three possibilities in Y108A, yielding almost the same result for conformations *C* and *P* and a 3 kcal/mol higher value for conformation *B*.

Since the geometry in which Tyr 106 blocks the binding pocket is energetically favorable, the correct settlement of this side chain in the native protein must be guaranteed during the folding process. However, in the Y108A mutant an alternative placement of the Tyr 106 ring becomes possible, as explained above. Once in conformation *B*, transitions into the binding pocket may easily occur as also observed during a 120 ps MD simulation (data not shown). According to the foregoing discussion, the binding observed for Fucα2Galβ4Glc may be attributed to the presence of some lectin molecules containing a "correctly" folded Tyr 106 side chain whereas the majority has the binding pocket blocked.

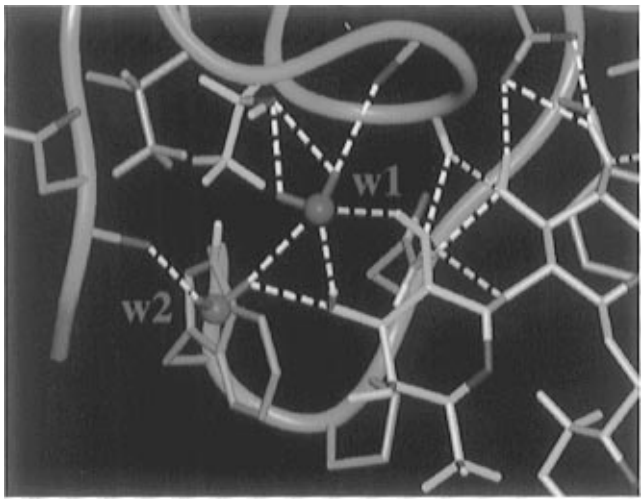

FIGURE 8: Partial view of the model of the ECorL Y108T−Fucα2Galβ4GlcNAcβ3 complex, illustrating the proposed role of water molecules (w1 and w2, shown in magenta) as bridges between the protein and the fucose residue (see text). The protein is in blue with oxygens and polar hydrogens marked in red and light gray, respectively. The Thr 108 side chain is colored in yellow, and the sugar residues are in light green with ring oxygens marked in red.

(D) Y108T. The binding properties of this mutant are similar to those of the native lectin, as shown in Figure 6. Threonine at a position equivalent to 108 in ECorL is also found in PNA, which in this respect differs from all the other sequences that were analyzed, as commented above. The crystal structure of this lectin, displaying only the α carbon trace, shows no deviations from other legume lectin structures. Furthermore, the effect described above involving Tyr 106 cannot occur when Tyr 108 is substituted by Thr. Although threonine is just slightly larger than alanine, this difference is enough to prevent Tyr 106 from assuming conformation *B* and thus block the binding site.

There is only one possibility to accommodate the side chain of Thr 108 favorably in the Y108T mutant without modifying the backbone geometry. In this conformation, the methyl group of Thr 108 is placed in the vicinity of Ile 150, whereas the hydroxyl group points to an open space occupied by Tyr in the native protein (Figure 8). In order to explore the role of possible solvent-mediated interactions for this mutant, a 100 ps MD simulation was run for the Y108T−Fucα2Galβ4GlcNAcβ3 trisaccharide complex in a spherical water shell ($R = 15$ Å) constructed around the ligand. Figure 8 shows part of the hydrogen bonds most frequently observed during the dynamics animation after a few picoseconds of equilibration. The protein contacts with the *N*-acetyllactosamine moiety were essentially the same as those observed in the complex between Fucα2Galβ4GlcNAcβ3 and ECorL (Figure 2), but other interactions with fucose appeared, involving the hydroxyl group of Thr 108 and two water molecules (Figure 8). The most internal of these water molecules (wl) is firmly retained at its position by a double hydrogen bond to the hydroxyl oxygen of Thr 108 and

hydrogen bonds to the carbonyl oxygen of Gly 105, the 2-OH and the 3-OH of fucose. A second water molecule (w2), bound to wl, serves as a bridge between 3-OH and the backbone oxygen of Asn 113. The simulation yields a probable explanation for the retained binding affinity of this mutant for the H5 type 2 glycosphingolipid as being due to water molecules serving as bridges between the terminal fucose of this glycosphingolipid and the protein, thus providing interactions compensating for those lost upon substitution of Tyr 108 by Thr.

Concluding Remarks. The above modeling results point to complex conformational dependencies between several residues around the binding site, and validation of some of the modeling results will require determination of high-resolution crystal structures for the recombinant forms in question. A particular effort was made in elucidating the role of Gln 219 in the binding to *N*-acetyllactosamine-containing glycosphingolipids. The conformational flexibility shown by the Gln 219 side chain, rather than being disadvantageous, appears to be very convenient since several favorable interactions, enhancing the affinity of the *N*-acetyllactosamine unit, may be accomplished for most of the possible conformations. On the other hand, Gln 219 is the only amino acid that interacts with the GlcNAc residue, while closing a sector of the binding site, which has the additional effect of restricting the movements of the ligand.

Particularly significant is the finding that ECorL exhibits a higher affinity for Fucα2Galβ4GlcNAcβ-terminated glycosphingolipids than for Galβ4GlcNAcβ-terminated ones, in contrast to ECL whose amino acid sequence thus may be assumed to differ from that of ECorL around the fucose part of the binding pocket. Furthermore, the perfect fit of fucose in the ECorL binding pocket argues for a redefinition of the binding specificity of this lectin, which thus should be considered Fucα2Galβ4GlcNAcβ3 specific in the first place. In practical applications which utilize the ECorL lectin, consideration of this binding specificity is very important since terminal Fucα2Galβ4GlcNAcβ3 is found in naturally occurring glycoconjugates. Whether this binding specificity has some relation with the biological function of this lectin remains to be investigated.

REFERENCES

Adar, R., & Sharon, N. (1996) *Eur. J. Biochem. 239*, 668−674.

Aggarwal, B. B., Eessalu, T. E., & Hass, P. E. (1985) *Nature 318*, 665−667.

Arango, R., Rozenblatt, S., & Sharon, N. (1990) *FEBS Lett. 264*, 109−111.

Arango, R., Adar, R., Rozenblatt, S., & Sharon, N. (1992) *Eur. J. Biochem. 205*, 575−581.

Arango, R., Rodriguez-Arango, E., Adar, R., Belenky, D., Loontiens, F., Rozenblatt, S., & Sharon, N. (1993) *FEBS Lett. 330*, 133−136.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 111*, 535−542.

Bourne, Y., Abergel, C., Cambillau, C., Frey, M., Rougé, P., & Fontecilla-Camps, J. C. (1990) *J. Mol. Biol. 214*, 571−584.

Delbaere, L. T. J., Vandonselaar, M., Prasad, L., Wilson, J., Wilson, K. S., & Dauter, Z. (1993) *J. Mol. Biol. 230*, 950−965.

Dessen, A., Gupta, D., Sabesan, S., Brewer, C. F., & Sachettini, J. C. (1995) *Biochemistry 34*, 4933−4942.

Einspahr, H., Parks, E. H., Suguna, K., Subramanian, E., & Suddath, F. L. (1986) *J. Biol. Chem. 261*, 16518−16527.

Hansson, G. C., Karlsson, K.-A., Larson, G., McKibbin, J. M., Strömberg, N., & Thurin, J. (1983) *Biochim. Biophys. Acta 750*, 214−216.

Imberty, A., Gerber, S., Tran, V., & Perez, S. (1990) *Glycoconjugate J. 7*, 27−54.

Imberty, A., Mikros, E., Koca, J., Mollicone, R., Oriol, R., & Perez, S. (1995) *Glycoconjugate J. 12*, 331−349.

Karlsson, K.-A. (1987) *Methods Enzymol. 138*, 212−220.

Keith, D. D., Tortora, J. A., & Yang, R. (1978) *J. Org. Chem. 43*, 3711−3713.

Laine, R. A., Stellner, K., & Hakomori, S.-i. (1974) *Methods Membr. Biol. 2*, 205−244.

Lis, H., Joubert, F. J., & Sharon, N. (1985) *Phytochemistry 24*, 2803−2809.

Loris, R., Steyaert, J., Maes, D., Lisgarten, J., Pickersgill, R., & Wyns, L. (1993) *Biochemistry 32*, 8772−8781.

Rini, J. M. (1995) *Annu. Rev. Biophys. Biomol. Struct. 24*, 551−577.

Shaanan, B., Lis, H., & Sharon, N. (1991) *Science 254*, 862−866.

Sharon, N., & Lis, H. (1989) in *Lectins*, pp 37−46, Chapman and Hall Ltd., London.

Sharon, N., & Lis, H. (1990) *FASEB J. 4*, 3198−3208.

Strömberg, N. (1988) *Thesis*, Göteborg University, Sweden.

Teneberg, S., Ångström, J., Jovall, P.-Å., & Karlsson, K.-A. (1994) *J. Biol. Chem. 269*, 8554−8563.

Weisgerber, S., & Helliwell, J. R. (1993) *J. Chem. Soc., Faraday Trans. 89*, 2667−2675.

Young, N. M., & Oomen, R. P. (1992) *J. Mol. Biol. 228*, 924−934.

BI962231H