

Structures of the Lectin IV of *Griffonia simplicifolia* and its Complex with the Lewis b Human Blood Group Determinant at 2.0 Å Resolution

Louis T. J. Delbaere†, Margaret Vandonselaar, Lata Prasad

Department of Biochemistry, University of Saskatchewan
Saskatoon, Saskatchewan, Canada S7N 0W0

J. Wilson Quail

Department of Chemistry, University of Saskatchewan
Saskatoon, Saskatchewan, Canada S7N 0W0

Keith S. Wilson and Zbigniew Dauter

European Molecular Biology Laboratory, Hamburg Outstation
Notkestrasse 85, D-2000, Hamburg 52, Germany

(Received 10 August 1992; accepted 20 October 1992)

The structures of the fourth lectin isolated from *Griffonia simplicifolia* (GS4) and its complex with the methyl-glycoside of the Lewis b human blood group determinant (Le^b-OMe) are reported at high resolution. The native GS4 crystal is isomorphous with the complexed GS4 crystal. The space group is $P4_22_12$ with unit cell dimensions $a = 78.9$ Å, $c = 89.1$ Å with one subunit of the lectin (bound to 1 Le^b-OMe in the complex) in the crystallographic asymmetric unit. The native GS4 structure was solved by the molecular replacement technique and least-squares refined (PROLSQ and X-PLOR). The orientation of the Le^b-OMe tetrasaccharide in the complex was established from a 2.8 Å difference map with coefficients ($F_{\text{complex}} - F_{\text{native}}$) and calculated phase angles from the native model. Both the final native and complex GS4 models consist of 1904 protein non-hydrogen atoms, one sulfate ion, one Ca ion, one Mn ion and three covalently-bound sugar residues N-linked to Asn18. In addition, the complex model has 47 Le^b-OMe non-hydrogen atoms. The two structures have 135 water molecules in common in addition to eight and nine unique water molecules in the native and complex structures, respectively. The root-mean-square deviations from ideal bond distances and angles are 0.016 Å, 3.2° and 0.016 Å, 3.0°, for the native and complexed GS4, respectively. The *R* index for all unique data from 8 to 2.0 Å is 0.187 for the native (19,204 reflections) and 0.181 for the complex (19,212 reflections). The tertiary structure of each subunit is similar to that of other leguminous lectins but the quaternary structure of the molecular dimer is different from that of any other lectin reported to date. The co-ordination about the Ca ion is pentagonal bipyramidal (with 1 long Ca²⁺-oxygen bond) and the co-ordination about the Mn ion is octahedral. Two conserved residues (Asp149 and Ser155) appear to be important because they are hydrogen-bonded to each other and to groups that co-ordinate the Mn ion. There are three *cis*-peptides in the polypeptide chain; two involve non-proline residues, one of which is homologous with other leguminous lectins and the other is unique to GS4. The two non-proline *cis*-peptides are located in the carbohydrate-binding site and are important for the specificity of the lectin.

The molecular recognition of Le^b-OMe by GS4 involves both polar and extensive non-polar interactions. Some key polar interactions of GS4 involve the side-chains of Asp89 and Asn135 and the peptide NH of Gly107, which are near the center of the shallow depression

† Author to whom all correspondence should be addressed.

which forms the carbohydrate-binding site. Several non-polar protein-carbohydrate interactions include six aromatic amino acid residues of GS4, namely Tyr105, Phe108, His114, Trp133, Trp138 and Tyr223 which make up a large portion of the carbohydrate-binding site. Several amino acid residues in the carbohydrate-binding site adjust positions by approximately 1 to 2 Å in order to optimize the hydrophilic and hydrophobic interactions in the complex; these rearrangements are similar in magnitude to those found to occur in antibody-antigen complexes. These results suggest that common aspects of receptor sites in proteins may be the occurrence of a significant number of aromatic residues and adjustments of <2 Å in residues upon formation of a complex.

Keywords: lectin; structure; carbohydrate-binding; crystallography; recognition

1. Introduction

Lectins are proteins that recognize and bind complex carbohydrate structures. This affinity is an example of the process of molecular recognition which is fundamental to all living processes. The fourth lectin isolated from the West-Central African shrub *Griffonia simplicifolia* (GS4†) is of special interest because of its strong non-covalent binding affinity for the Lewis b (Le^b) and Y human blood group determinants, α -L-Fuc(1 → 2) β -D-Gal(1 → 3)[α -L-Fuc(1 → 4)] β -D-GlcNAc and α -L-Fuc(1 → 2) β -D-Gal(1 → 4)[α -L-Fuc(1 → 3)] β -D-GlcNAc, respectively (Shibata *et al.*, 1982). GS4 binds similar surface topologies in the Lewis b and Y human blood group determinants (Spohr *et al.*, 1985). GS4 is a molecular dimer of M_r 56,000 with non-covalently-linked subunits. Each subunit consists of 243 amino acid residues, N-linked carbohydrate, one Mn ion, one Ca ion and one carbohydrate-binding site. The amino acid sequence of GS4 is more than 40% identical to that of favin, concanavalin A (ConA) and pea lectin (L. B. Smillie, personal communication). Heterogeneity has been detected at five amino acid residue positions of the polypeptide chains (Nikrad *et al.*, 1989). Both subunits have carbohydrate covalently bound, one has carbohydrate attached to both Asn5 and Asn18 whereas the other has carbohydrate only on Asn18 (Nikrad *et al.*, 1990).

Several analogs of Le^b -OMe (the methyl-glycoside of Le^b) have been synthesized and the thermodynamic parameters for their binding to GS4 have been determined (Spohr *et al.*, 1986). Complexes of both Le^b -OMe and analogs of Le^b -OMe have crystallized isomorphously with the native GS4 (Vandonselaar *et al.*, 1987). An extensive structural study of the native lectin and these complexes is underway in order to determine the detailed interactions involved in this molecular recognition, to determine the role of water molecules in this binding and if possible, to correlate the structural results with the thermodynamic parameters. The 2.8 Å (1 Å = 0.1 nm) analysis of the complex of GS4 with Le^b -OMe has been reported (Delbaere *et al.*, 1990).

This analysis and that of the native GS4, have been extended to 2.0 Å resolution; this results in the determination of the atomic positions to a much greater accuracy, allows the location of some N-linked carbohydrate and permits the location of several well-ordered water molecules.

2. Materials and Methods

Native GS4 and synthetic Le^b -OMe were prepared (Spohr *et al.*, 1985) and kindly provided by U. Spohr and R. U. Lemieux. The complex (prepared by mixing lectin and Le^b -OMe in a molar ratio of 1 : 2.2) and native GS4 were crystallized under the same conditions: 36% (w/v) saturated ammonium sulfate solution with 0.1 M-Pipes buffer at pH 6.2 (Vandonselaar *et al.*, 1987). The space group is $P4_22_12$ with $a = 78.9$ Å, $c = 89.1$ Å with 1 subunit of the lectin in the crystallographic asymmetric unit. There is 1 Le^b -OMe/subunit in the complex. The structure of native GS4 was solved by molecular replacement with the co-ordinates of 1 subunit of pea lectin (Einspahr *et al.*, 1986) using 2.8 Å data collected by an Enraf-Nonius CAD4F 4-circle diffractometer (Delbaere *et al.*, 1989); 1 subunit of favin (Reeke & Becker, 1986) gave the same result. Atomic co-ordinates for the Le^b -OMe tetrasaccharide were obtained from hard sphere exo-anomeric (HSEA) calculations (Lemieux *et al.*, 1980) and kindly provided by H. Beierbeck. The location of the carbohydrate-binding site of GS4 was resolved by use of an iodo-derivative of Lewis b which had the 6-hydroxyl group of the GlcNAc moiety replaced by an iodine atom (Spohr & Lemieux, 1988); a 6 Å difference Patterson map calculated with coefficients from the structure amplitudes of the complex of GS4 with iodo- Le^b -OMe and the complex of GS4 with Le^b -OMe was used. The approximate orientation of the Le^b -OMe tetrasaccharide was established from a 2.8 Å difference map with coefficients derived from the structure amplitudes of the complex and the native GS4 crystals, and phases calculated from the native GS4 structure (Delbaere *et al.*, 1989).

Subsequently, data were collected to 2.0 Å at the EMBL Outstation, DESY, Hamburg, Germany. The native GS4 data were obtained on the X11 source with the use of an image plate scanner. The synchrotron data were merged with CAD4F data from 5 native GS4 crystals to give a merging R value of 0.10 and 19,204 unique reflections within the range of 8.0 to 2.0 Å d spacings. Data from crystals of the complex of GS4 with Le^b -OMe were obtained on the X11 source with photographic film; the synchrotron data were merged with CAD4F data from 5 complex crystals to give a merging R value of 0.12 and 19,212 unique reflections within the range of 8.0 to 2.0 Å d spacings. Data reduction was carried out with a modified

† Abbreviations used: GS4, fourth lectin isolated from *Griffonia simplicifolia*; Le^b , Lewis b; ConA, concanavalin A; HSEA, hard sphere exo-anomeric; LOLI, *Lathyrus ochrus* solectin I; EcorL, *Erythrina corallodendron* lectin.

Table 1
Final refinement parameters for native and complexed GS4

	Native GS4	Complex
<i>R</i> -index	0.187	0.181
r.m.s. deviation in bond distances (Å)	0.016	0.016
r.m.s. deviation in angles (deg.)	3.21	3.02
No. of solvents	144	145
No. of reflections (% completeness)	19,204 (99.5)	19,212 (99.6)
Mean temperature factors (Å ²):		
Backbone	26.0	25.6
Side-chains	27.9	27.2
Asn18-linked sugar	41.1	38.6
Water molecules	37.0	41.3
Average occupancy of water molecules	0.87	0.92

All unique measured data (8.0 to 2.0 Å) were used in the least-squares refinements; there was no σ cutoff.

version of the MOSFLM program suite (CCP4, 1979); reflection intensities were submitted to a Bayesian treatment (French & Wilson, 1978). Molecular replacement computations were carried out with the CCP4 suite of programs (Leslie *et al.*, 1986).

Initial refinement and model adjustment were carried out with a version of the program PROLSQ (Konnert & Hendrickson, 1980) which had been modified to take into account intermolecular contacts (Vincent & Priestle, 1985). Subsequently the molecular dynamics program with simulated annealing, X-PLOR (Brünger, 1988), was used for least-squares refinement. This latter refinement typically consisted of a prestage of 160 cycles of minimization, followed by slow cooling from an initial temperature of 4000 K, decreasing in steps of 50 K with a timestep of 0.5 fsec, then another 120 cycles of minimization, and finally, 60 cycles of individual isotropic *B* factor refinement. A modification had been made in the protein parameter file to allow peptide bonds to refine as either *cis* or *trans*. Harmonic restraints were applied to solvent molecules during the slow cooling stage (harmonic force constant = 10.0 kcal/(mol Å²)). A modification had also been added to define the N-terminal residue as a pyrrolidone carboxylate instead of a glutamate (Nikrad *et al.*, 1989). The methoxy group at C-1 of Le^b-OMe was defined as part of the GlcNAc residue in the parameter file.

Since the subunits are not identical, initial refinements were carried out in the less symmetrical space group, *P*₄₂, which has 1 dimer in the crystallographic asymmetric unit. The residues at which heterogeneity had been detected (Nikrad *et al.*, 1989) were designated as alanines. As refinement and model building proceeded and no differences between the subunits could be detected in the electron density maps, space group *P*₄₂₁₂ was assumed to be correct and was used for the final stages of refinement. The coefficients used for the electron density maps were initially $3|F_o| - 2|F_c|$ and the phase angles were the calculated ones for the current model. Manual model building was carried out using the program FRODO (Jones, 1978; Oatley, 1985) on a Silicon Graphics IRIS 3030 workstation and TURBO-FRODO (Roussel & Cambillau, 1992) on a Silicon Graphics IRIS 4D30TG workstation. Water molecules were located with the use of difference maps calculated first at 2.15 Å and later at 2.0 Å resolution. During the final stages of least-squares refinement the slow cooling stage was omitted and only conventional positional refinement and individual isotropic *B* factor refinement were carried out. Both $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ coefficients along with calculated

phases were used for the maps. PROLSQ was used to refine the occupancies of the solvent molecules; indicated shifts on *B* values and occupancies were applied in alternate cycles. Water molecules that refined to very high temperature factors (>75 Å²) or very low occupancy values (<0.30) were eliminated. The occupancies of the water molecules were fixed in X-PLOR for the final cycles of refinement. Final refinement parameters are given in Table 1 and the atomic co-ordinates as well as the structure factors are being deposited with the Protein Data Bank (Bernstein *et al.*, 1977), accession numbers 1LEC and 1LED for the native and complex, respectively.

3. Results and Discussion

The variation in the *R*-factor versus $1/d$ is given in Figure 1 for native GS4 and for the complex. Luzzatti (1952) plots suggest that the expected coordinate error is between 0.2 and 0.25 Å for both crystal structures. The variation of the average isotropic temperature factor along the polypeptide chain for both the main-chain and the side-chain atoms of native GS4 is shown in Figure 2. The Ramakrishnan & Ramachandran (1965) plot is shown in Figure 3. All but two of the ϕ, ψ values fall within the allowed regions; one of these corresponds to Asn242 which has very high temperature factors

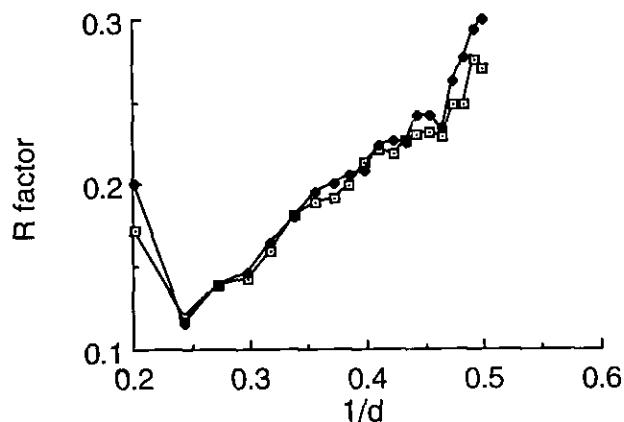


Figure 1. Plot of *R*-factor versus $1/d$ spacing for (—●—) native GS4 and (—□—) the complex.

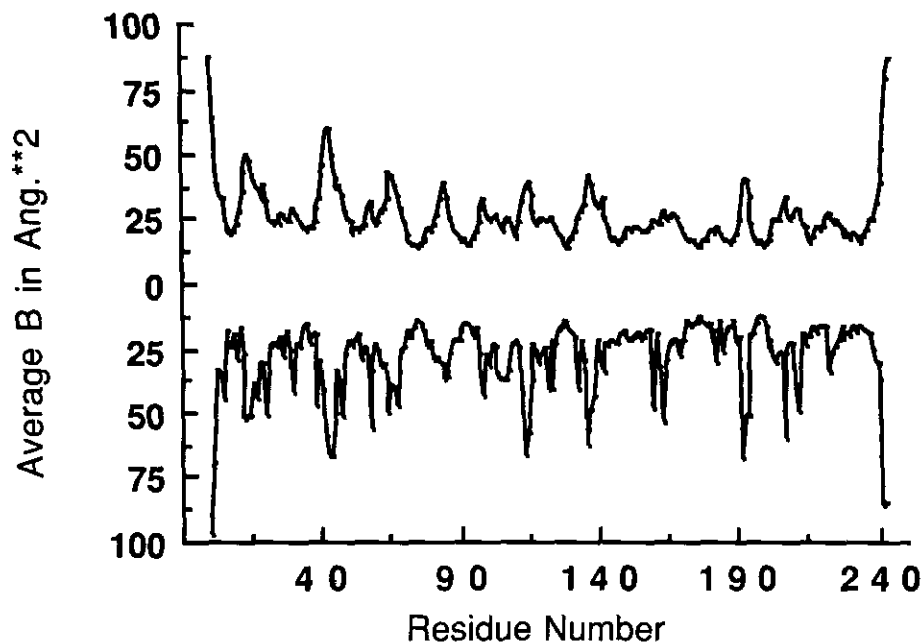


Figure 2. Variation of mean isotropic temperature factor of native GS4 for main-chain atoms of each residue (upper plot) and variation of mean isotropic temperature factor for side-chain atoms of each residue (lower plot).

(see Fig. 2). The second anomalous ϕ, ψ value corresponds to Lys103 and there is no obvious reason for this deviation.

(a) *Secondary and tertiary structure*

The overall folding of the polypeptide chain of the GS4 monomer is illustrated in stereo in Figure 4. The monomer consists of two major antiparallel β -sheets of secondary structure, one of six strands and the other of seven strands, similar to the structures of pea lectin (Einspahr *et al.*, 1986), favin (Reeke & Becker, 1986), ConA (Hardman & Ainsworth, 1972; Becker *et al.*, 1975), *Lathyrus ochrus* isolectin I (LOLI) (Bourne *et al.*, 1990a) and *Erythrina corallodendron* lectin (EcorL; Shaanan *et al.*, 1991). The reverse turns which occur in the polypeptide chain of GS4 are listed in Table 2, in categories of standard types of turns defined by the ϕ, ψ values of the inner two residues of a four residue turn (Chou & Fasman, 1977); a β -turn is considered to be a standard type if the ϕ, ψ values are within 50° of ideal values. GS4 contains one turn of α -helix (Glu132 to Tyr167) which is not found in the other leguminous lectins; this turn is in a region that has a three residue insertion in GS4 (L. B. Smillie, personal communication).

(b) *Quaternary structure*

In space group $P4_22_12$ there are two unique crystallographic 2-fold axes which relate subunits of GS4, one parallel to the c -axis and one along the (001) face diagonal. The latter 2-fold axis is assumed to generate the GS4 molecular dimer because there are more close contacts between subunits across this axis than across the former axis.

The quaternary structure of GS4 is different from that of any other lectin that has been reported to date. A stereo diagram which illustrates the molecular dimer of GS4 is shown in Figure 5, along with the pea lectin dimer for comparison; one subunit of GS4 and one subunit of pea lectin were

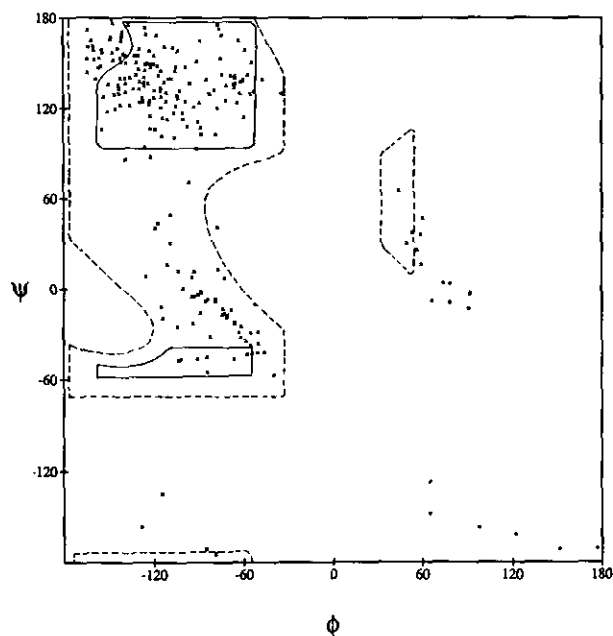


Figure 3. The polypeptide ϕ, ψ plot of the refined structure of native GS4. The circles denote glycine residues. The continuous curves enclose areas that are the fully-allowed conformational regions for $\tau(C^*)$ of 110° and the broken curves encircle the area of acceptable van der Waals contacts for $\tau(C^*)$ of 115° (Ramakrishnan & Ramachandran, 1965).

Table 2
 ϕ, ψ angles (deg.) in turns of native GS4

	ϕ_1	ψ_1	ϕ_2	ψ_2
Standard type I	-60	-30	-90	0
Residues 42-43	-50	-28	-92	11
65-66	-50	-36	-73	7
98-99	-59	-34	-79	-6
107-108	-52	-10	-86	-8
114-115	-51	-42	-71	-18
117-118	-55	-37	-94	7
120-121	-61	-24	-74	-16
137-138	-54	-42	-104	12
181-182	-85	-45	-93	-4
192-193	-40	-57	-117	43
Standard type I'	60	30	90	0
Residues 109-110	56	26	74	4
Standard type II	-60	120	80	0
Residues 18-19	-52	131	91	-3
32-33	-59	114	91	-2
239-240	-34	129	78	-8
Standard type II'	60	-120	-80	0
Residues 26-27	65	-127	-93	-11
106-107	64	-148	-52	-10
Standard type III	-60	-30	-60	-30
Residues 20-21	-84	-54	-104	-47
180-181	-74	-12	-85	-45
Type VI, a <i>cis</i> -Pro at position 2				
Residues 85-86	-142	135	-77	175

	ϕ	ψ
Standard α -helix	-57	-47
Residue 162	-60	138
Residue 163	-46	-42
Residue 164	-71	-17
Residue 165	-91	-46
Residue 166	-58	-43
Residue 167	-111	16
Residue 205	-62	-32
Residue 206	-63	-29
Residue 207	-89	-15
Residue 208	-114	-19
Residue 209	-119	154

overlapped by use of the program LSQKAB (CCP4, 1979). Favin, the ConA dimer and LOLI all have a similar quaternary structure to pea lectin. EcorL cannot form the standard dimer because of potential steric interference from the N-glycosylated heptasaccharide on Asn17 (Shaanan *et al.*, 1991).

GS4 also does not form the canonical dimer, possibly because this would bury a charged residue on each subunit, Glu58, at the subunit interface; the corresponding topologically-equivalent residues are Leu49 in favin, Ser48 in pea lectin, Ser48 in LOLI and Ala177 in ConA. The GS4 molecular dimer is

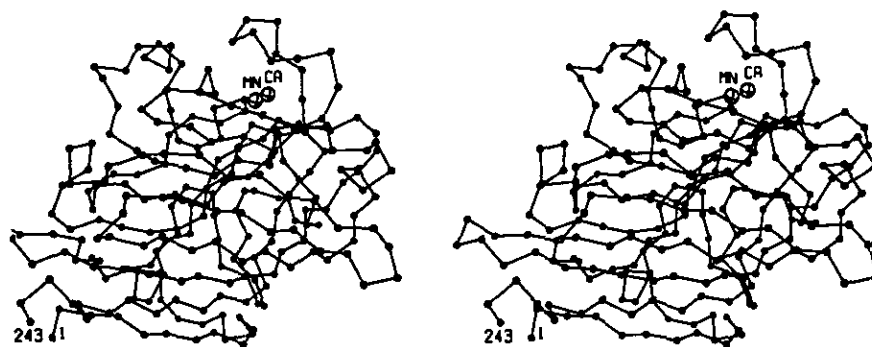


Figure 4. Stereoscopic view of the polypeptide chain of the crystallographically-unique subunit of native GS4; only the α -carbon atoms of the amino acid residues are illustrated. The N- and C-terminal residues are labeled 1 and 243, respectively. The positions of the Ca and Mn ions are labeled CA and MN, respectively; in this orientation the carbohydrate-binding site is located at the top of the subunit near the Ca ion.

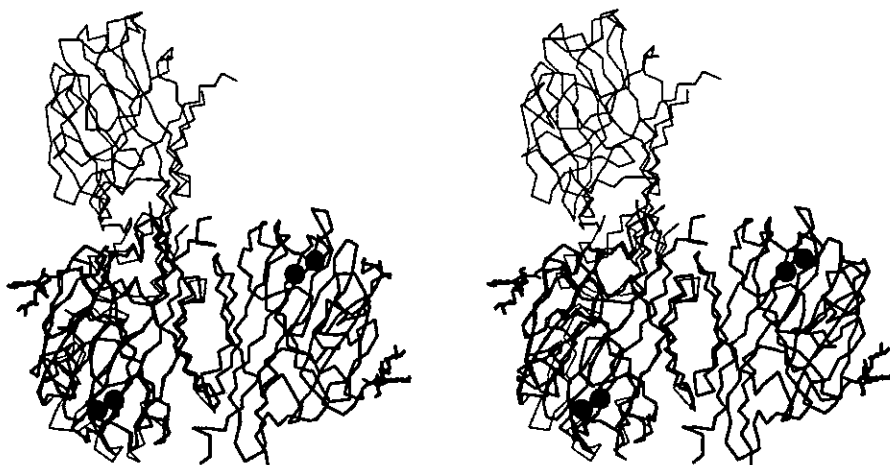


Figure 5. The quaternary structure of native GS4 (thick lines) and pea lectin (thin lines). On each GS4 subunit, 3 sugar residues (which are N-linked to Asn18) have been located in the crystal structure and are also illustrated. The GS4 molecular dimer is formed by the interaction of the 6-stranded β -sheets of each subunit, in a manner that is unique to GS4.

formed by association of the six-stranded β -sheets of each monomer, with the β -strands of one sheet being almost perpendicular to those of the other (see Fig. 5). There is a surface area of 900 \AA^2 per GS4 subunit buried at the interface (calculated with a probe radius of 1.4 \AA ; Connolly, 1983); this value is similar to the value occurring in favin and pea lectin, but is greater than the 700 \AA^2 per subunit buried at the interface of EcorL.

Both polar and non-polar interactions are involved in dimer formation of GS4. The hydrogen-bonding interactions which occur between the subunits of the molecular dimer are shown in Table 3; note that due to the 2-fold crystallographic symmetry, each interaction listed in Table 3 (except for that of Thr186 OG1) occurs twice in the molecular dimer. Thus, there are 13 hydrogen bonds directly between the side-chains of each subunit and ten hydrogen-bonding interactions *via* water molecules. In addition, there are 139 van der Waals contacts ($< 4 \text{ \AA}$) between atoms of the two subunits.

(c) *GS4 molecular dimer complex with Le^b-OMe*

The GS4 molecular dimer complex is illustrated in Figure 6 with a Le^b-OMe bound to each subunit of GS4. The bonds between the non-hydrogen atoms of the Le^b-OMe tetrasaccharides are in gold; the carbohydrate-binding sites are located at the extremes of the molecule. Upon complex formation, the buried surface area for each subunit (calculated with a probe radius of 1.4 \AA ; Connolly, 1983) is 265 \AA^2 on GS4 and 237 \AA^2 on Le^b-OMe.

(d) *N-linked carbohydrate*

The oligosaccharide which has been shown to be covalently bound to Asn5 of only one subunit of the molecular dimer (Nikrad *et al.*, 1990) is likely located near the molecular dimer interface; however there is no meaningful electron density for this

carbohydrate in the maps, probably due to the 2-fold crystallographic averaging over the same site of both Asn5 residues in the molecular dimer and also possibly due to the mobility of this carbohydrate. Three sugar residues of the Asn18 N-linked heptasaccharide have been located and are shown in Figure 5. The electron density, illustrated in stereo in Figure 7, for an α -L-Fuc(1 \rightarrow 3) and a β -D-GlcNAc(1 \rightarrow 4) linked to the β -D-GlcNAc-Asn18 is continuous and strong but there is no significant electron density for the remaining four sugar residues which are part of the carbohydrate bound to Asn18 (Nikrad *et al.*, 1990), probably due to the mobility of these remaining sugar residues. There are two hydrogen bonds between this covalently-linked carbohydrate and the protein portion of GS4; the NH of the N-acetyl group of the GlcNAc directly linked to Asn18 is hydrogen-bonded to the carbonyl oxygen atom of Phe24; the carbonyl group of the N-acetyl moiety of the second GlcNAc is hydrogen-bonded to Ser101 OG.

Table 3

Hydrogen-bonding interactions ($< 3.3 \text{ \AA}$) between subunits in native GS4 molecular dimer

Thr159 OG1	Arg181 *NH1		
Asp196 OD1	His233 *NE2		
Asp179 OD2	Ser200 *N		
Arg181 NH2	Ile198 *O		
Ser182 OG	Ser200 *O		
Thr186 OG1	Thr186 *OG1		
Arg195 NH1	Asp239 *OD1		
Ser71 OG	water383	Ile198 *O
Thr159 OG1	water315	Asp239 *OD1
Arg160 O	water315	Asp239 *OD1
Thr177 OG1	water301	Ser200 *OG
Arg181 NH1	water383	Ile198 *O

The atoms labeled with * refer to crystallographically 2-fold equivalent atoms in the other subunit.

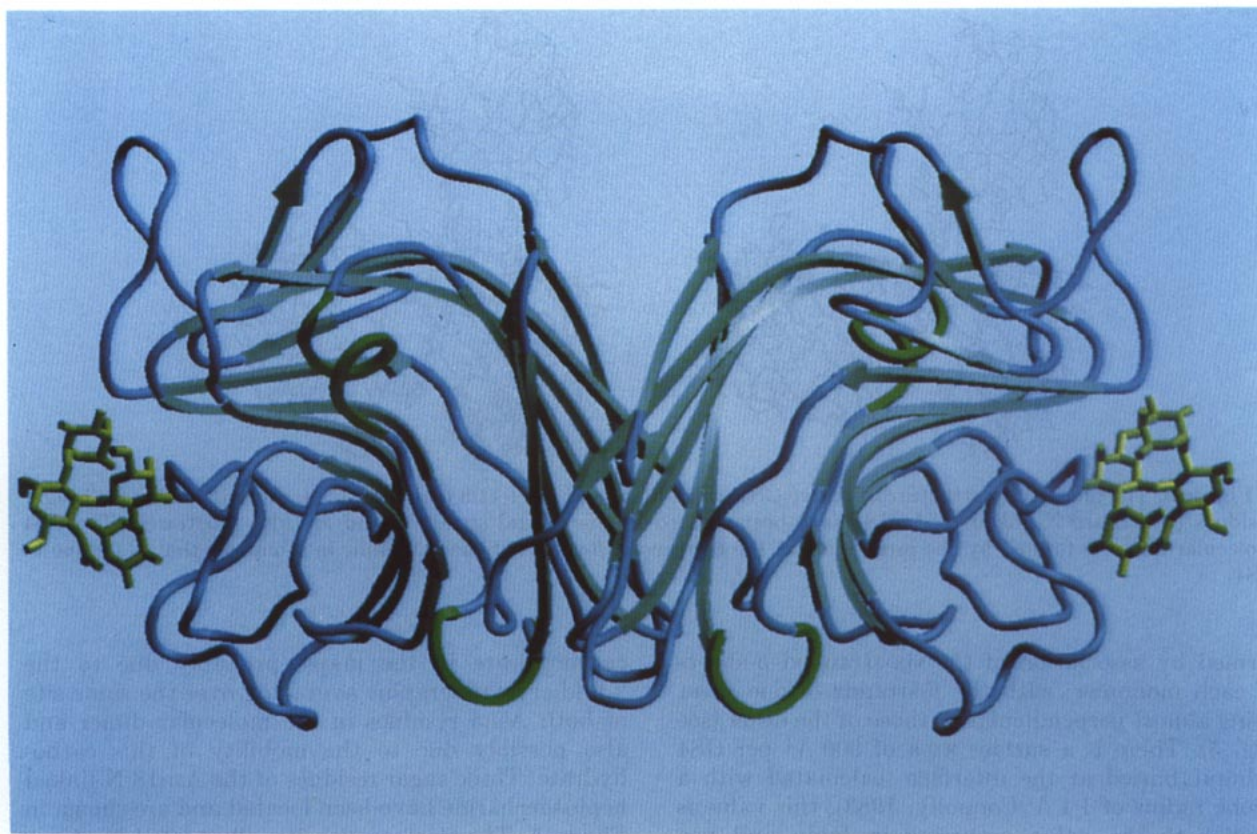


Figure 6. Molecular dimer of the complex of GS4 and Le^b-OMe. The strands of the β -sheets are illustrated as blue-green ribbons with arrowheads at the C-terminal ends. The bonds between the non-hydrogen atoms of Le^b-OMe are gold. The Le^b-OMe binding sites are located at the extremities of GS4. This drawing was created with the use of the program SETOR (Evans, 1992).

(e) *Metal ions and the carbohydrate-binding site*

The presence of the Ca and Mn ions is necessary for the carbohydrate-binding activity of GS4 and related lectins. The side-chains of Asp131 and Asp140 bridge the Mn and Ca ions, which are 4.19 Å apart. The co-ordination around the Mn and Ca sites in native GS4 is illustrated in Figure 8 and the

bond lengths and angles around the metal ions are given in Table 4. The co-ordination distances to the metal ions were not restrained during refinement. The Ca²⁺ co-ordination consists of six oxygen atoms (from water molecules or carboxylate groups) in a distorted octahedral configuration (2.3 Å mean distance) plus a 2.7 Å interatomic distance to a

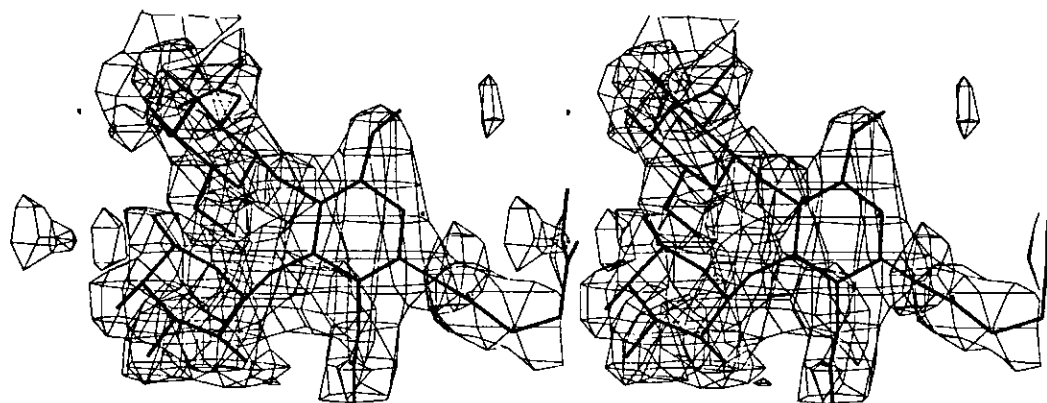


Figure 7. The electron density map in the region of Asn18 in native GS4. Note that there is electron density for the α -L-Fuc(1 \rightarrow 3) β -D-GlcNAc(1 \rightarrow 4) β -D-GlcNAc-Asn18 moiety but there is no electron density for the additional 4 sugar residues which are also part of the covalently-bound carbohydrate (Nikrad *et al.*, 1990). The coefficients for this map are $2|F_o| - |F_c|$ and the phase angles are those calculated from the final model; the electron density is contoured at 1.5 σ .

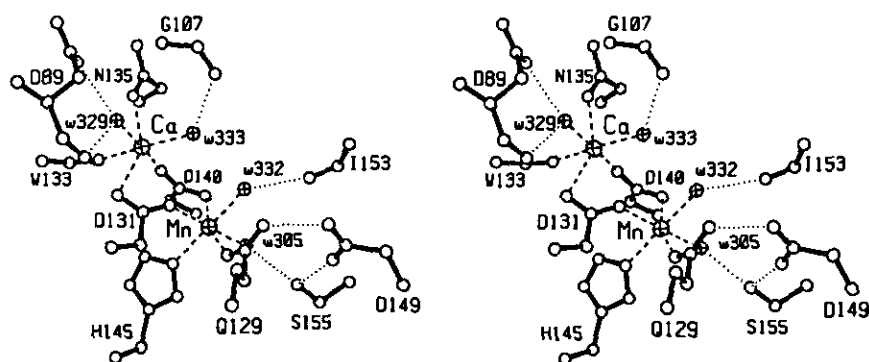


Figure 8. Stereo view of the metal-binding sites in native GS4. The α -carbon atoms and certain non-hydrogen atoms of relevant amino acid residues are illustrated.

Table 4
Bond lengths and angles for the Ca^{2+} and Mn^{2+}
binding sites in native GS4

Atom	Bond length (Å)
A. Distance to Ca^{2+}	
Asp13 OD1	2.4
Trp133 O	2.3
Asn135 OD1	2.4
Asp140 OD1	2.4
water329	2.2
water333	2.4
Asp131 OD2	2.7
Mn^{2+}	4.19
B. Distance to Mn^{2+}	
Gln129 OE1	2.2
Asp14 OD2	2.1
Asp14 OD2	2.3
His145 NE2	2.3
water305	2.0
water332	2.2
C. Bond angles (deg.)	
Asp13 OD1– Ca^{2+} –Asn135 OD1	147
Trp133 O– Ca^{2+} –water333	174
Asp140 OD1– Ca^{2+} –water329	171
Asp13 OD1– Ca^{2+} –Trp133 O	74
Asp13 OD1– Ca^{2+} –Asp140 OD1	101
Asp13 OD1– Ca^{2+} –water329	80
Asp131 OD1– Ca^{2+} –water333	109
Trp133 O– Ca^{2+} –Asn135 OD1	81
Trp133 O– Ca^{2+} –Asp140 OD1	73
Trp133 O– Ca^{2+} –water329	99
Asn135 OD1– Ca^{2+} –Asp140 OD1	91
Asn135 OD1– Ca^{2+} –water329	83
Asn135 OD1– Ca^{2+} –water333	98
Asp140 OD1– Ca^{2+} –water333	101
water329– Ca^{2+} –water333	87
Gln129 OE1– Mn^{2+} –Asp140 OD2	165
Asp131 OD2– Mn^{2+} –water305	178
His145 NE2– Mn^{2+} –water332	177
Gln129 OE1– Mn^{2+} –Asp131 OD2	91
Gln129 OE1– Mn^{2+} –His145 NE2	99
Gln129 OE1– Mn^{2+} –water305	90
Gln129 OE1– Mn^{2+} –water332	84
Asp131 OD2– Mn^{2+} –Asp140 OD2	99
Asp131 OD2– Mn^{2+} –His145 NE2	86
Asp131 OD2– Mn^{2+} –water332	95
Asp140 OD2– Mn^{2+} –His145 NE2	95
Asp140 OD2– Mn^{2+} –water305	80
Asp140 OD2– Mn^{2+} –water332	83
His145 NE2– Mn^{2+} –water305	96
water305– Mn^{2+} –water332	83

seventh oxygen atom, Asp131 OD2, which is also co-ordinated to the Mn ion (2.1 Å interatomic distance). The position of this seventh ligand completes a pentagonal bipyramidal co-ordination about the Ca ion. It has been suggested (Strynadka & James, 1989) that Ca ions in proteins prefer seven co-ordination, and this may be an example of such co-ordination. However, the Asp131 OD2 atom would be a poor co-ordinating group to the Ca ion because it is already strongly co-ordinated to the Mn ion. In addition, the co-ordination geometry is strained with the angle Asp131 CG–Asp131 OD2– Ca^{2+} at 87°, which is less than ideal, whereas the angle Asp131 CG–Asp131 OD2– Mn^{2+} is 101°. Although the geometry is not ideal for a co-ordinate bond, there is likely an electrostatic attraction between the positively-charged Ca ion and the residual negative charge on Asp131 OD2. Oxygen atoms acting as bridges between Mn ions are found in the Mn(acetylacetonate)₂ trimer (Shibata *et al.*, 1985) where singly-co-ordinated oxygen to Mn^{2+} distances are 2.1 Å and bridging oxygen to Mn^{2+} distances are in the range 2.16 to 2.25 Å. The larger radius (Shannon & Prewitt, 1969) of a Ca ion (1.00 Å for 6 co-ordination and 1.07 Å for 7 co-ordination) compared to 0.82 Å for Mn^{2+} would lead to an expected Ca^{2+} –oxygen distance of approximately 2.4 Å, 0.3 Å shorter than the observed value. Any adjustment of the model to give a shorter Ca^{2+} –Asp131 OD2 distance would increase the Mn^{2+} –Asp131 OD2 distance. Thus, there is bonding of a seventh ligand to the Ca^{2+} , but this seventh bond is much longer and weaker than the other six bonds.

The Mn ion is six co-ordinated with a mean co-ordination distance of 2.2 Å. The Mn ion is relatively close to ideal octahedral geometry (Table 4) with a mean angle of 174° between *trans*-co-ordinating atoms and a mean angle of 90° between adjacent-co-ordinating atoms. It should be noted that two amino acid residues, Asp149 and Ser155, hydrogen bond to groups which co-ordinate the Mn ion (Gln129 and water305, respectively) as well as hydrogen bond to each other; these amino acid residues are topologically conserved in other leguminous lectins and these interactions with important Mn^{2+} -co-ordinating groups provide a reason for their conservation.

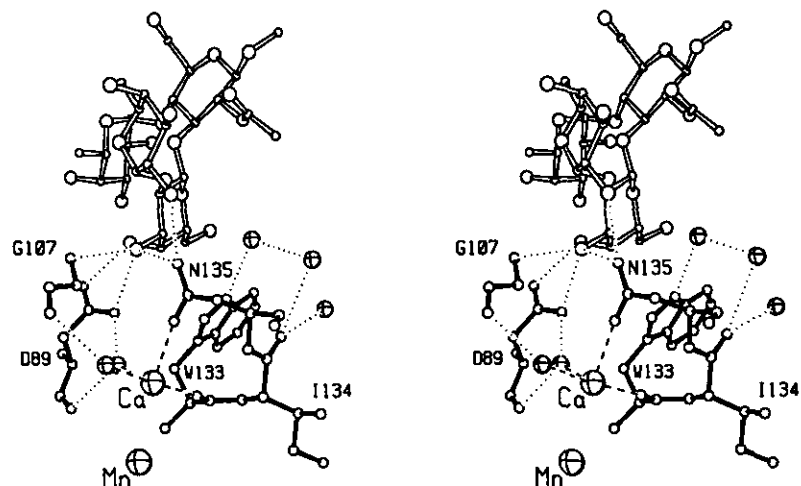


Figure 9. Carbohydrate-binding site of the complex in the region of the Ca and Mn ions (large cross-hatched spheres). Bonds between non-hydrogen atoms of GS4 are filled whereas those between non-hydrogen atoms of Le^b-OMe are open. Water molecules are illustrated as small cross-hatched spheres. Co-ordinate bonds to the metal ions are shown as broken lines; hydrogen bonds are represented by dotted lines.

Figure 9 illustrates the location of the Ca and Mn ions with respect to Le^b-OMe in the carbohydrate-binding site of GS4. The sugar residues of Le^b-OMe will be denoted as: **a**, β -D-GlcNAc; **b**, β -D-Gal; **c**, α -L-Fuc(1 \rightarrow 4) linked to β -D-GlcNAc; and **d**, α -L-Fuc(1 \rightarrow 2) linked to β -D-Gal in the following[†] discussion. As shown in Figure 9, the Ca ion, in particular, is very important for the correct positioning and orientation of certain GS4 subunit amino acid residues which interact with Le^b-OMe. A water molecule that co-ordinates the Ca²⁺ is hydrogen-bonded to the carbonyl group of Asp89 as well as to Asp89 OD2; the carboxylate group of Asp89 accepts hydrogen bonds from O3 and O4 of the **b**Gal. The second water molecule which co-ordinates the Ca²⁺ is hydrogen-bonded to the carbonyl group of Gly107; the peptide NH of Gly107 is, in turn, hydrogen-bonded to Asp89 OD2. The carbonyl group of the side-chain of Asn135 is co-ordinated to

the Ca²⁺ and the -NH₂ unit of this side-chain is hydrogen-bonded to O-3 of **b**Gal and to O-2 of **d**Fuc. The positive charge on the Ca²⁺ would serve to strengthen the above-mentioned hydrogen bonds between GS4 and Le^b-OMe. The carbonyl group of Trp133 is co-ordinated to Ca²⁺ and the side-chain of Trp133 makes extensive van der Waals contacts with **b**Gal. In summary, the metal ions *via* a network of interactions indirectly contribute to the proper location of important groups which interact with Le^b-OMe in the carbohydrate-binding site.

(f) *Cis-peptide bonds and the carbohydrate-binding site in native GS4*

There are three peptide bonds in each subunit of GS4 that have the *cis*-configuration, Ala85-Pro86, Ala88-Asp89 and Val221-Gly222. The stereo diagram in Figure 10 shows that not only are these

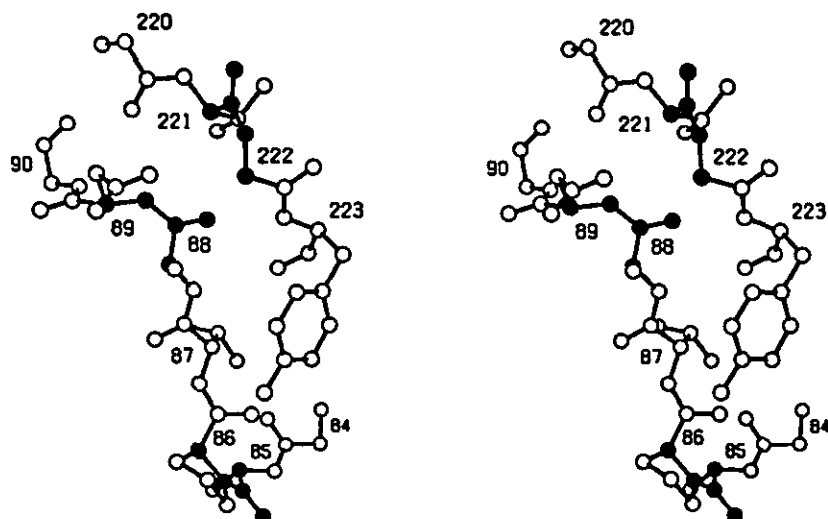


Figure 10. Stereo view of the non-hydrogen atoms of the residues in the region of the 3 *cis*-peptide bonds in native GS4. The *cis*-peptide bonds (filled circles) are between Ala85-Pro86, Ala88-Asp89 and Val221-Gly222.

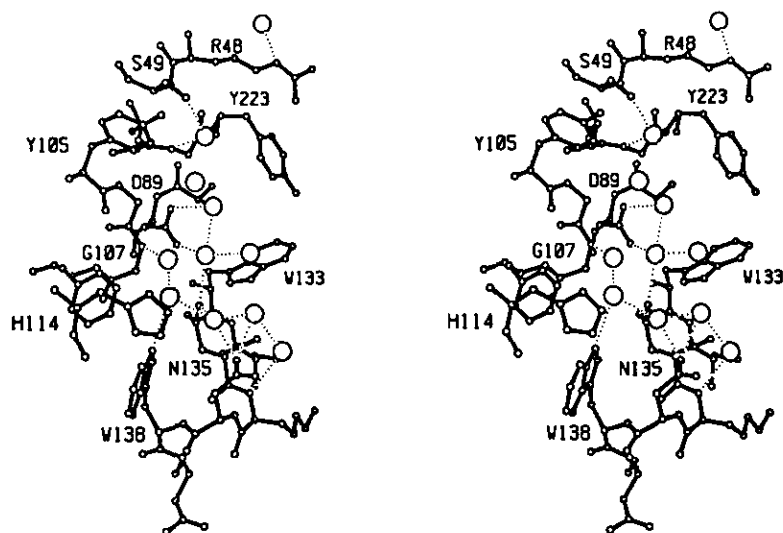


Figure 11. Stereo diagram of the carbohydrate-binding site in native GS4. Non-hydrogen atoms of relevant residues are illustrated; water molecules are denoted as large open circles. Hydrogen bonding is shown in dotted lines.

cis-peptides relatively close to each other in space, but the two non-proline peptides are actually next to each other on adjacent β -strands. The first *cis*-peptide involves a proline residue at a sharp bend in the polypeptide chain, which is not uncommon in proteins. The second *cis*-peptide, Ala88–Asp89, is structurally analogous to that found in other related lectins in which the side-chain of the aspartate residue is involved in binding carbohydrate to these lectins (Reeke & Becker, 1986; Rini *et al.*, 1989; Derewenda *et al.*, 1989; Bourne *et al.*, 1990b; Shaanan *et al.*, 1991). The third *cis*-peptide is unique to GS4 and is involved in the specificity of this lectin towards the Le^b tetrasaccharide. The amino acid residues and 11 water molecules in the region of the carbohydrate-binding site of native GS4 are illustrated in Figure 11; the hydrogen-bonding interactions are listed in Table 5. Water440 forms hydrogen bonds to both Gly222 N and Ser49 OG

Table 5
Hydrogen bonds (<3.3 Å) in the region of the carbohydrate-binding site of native GS4

Arg48 NE	water436
Ser49 OG	water440
Asp89 OD1	water438
Asp89 OD2	water437
Gly107 N	water439
Asn135 ND2	water382
Asn135 ND2	water437
Lys136 N	water409
Asp137 OD1	water382
Asp137 OD1	water420
Trp138 NE1	water442
Gly222 N	water440
water382	water420
water382	water442
water409	water420
water437	water438
water437	water443
water439	water442
water440	water441

while water437 and water438 hydrogen bond to the carboxylate group of Asp89, stabilizing the two *cis*-peptide configurations and helping to align the side-chains in the proper orientation to receive the carbohydrate.

(g) *Carbohydrate-binding site in the GS4 complex with Le^b-OMe*

A stereo diagram of the non-hydrogen atoms of Le^b-OMe, water molecules and amino acid residues in the carbohydrate-binding site of GS4 is shown in Figure 12. Hydrogen bonds are illustrated as dotted lines and are listed in Table 6. There are 12 hydrogen-bonding interactions that occur directly between GS4 and Le^b-OMe. In addition there are several hydrogen bonds in the carbohydrate-binding site between water molecules and atoms of GS4 or Le^b-OMe. Upon complexation, eight of the 11 water molecules in the area of the combining site (see Fig. 11) are displaced; the three water molecules which are common to both structures are denoted as cross-hatched spheres in Figure 12. Nine additional well-ordered water molecules are found in the carbohydrate-binding site in the complex and are illustrated as large open spheres in Figure 12. Thus, there is a net increase of one water molecule in the crystallographic asymmetric unit of the complex as compared to native GS4. The electron density corresponding to the Le^b-OMe is illustrated in Figure 13.

(h) *Key polar interactions*

Inhibition studies on complexes of GS4 with Le^b-OMe and several synthetic analogs of Le^b-OMe, indicated, that there are three hydroxyl groups, termed the key polar groups, which are essential for complex formation (Spohr *et al.*, 1985, 1986; Lemieux, 1989). These groups (the 3- and 4-hydroxyl groups of bGal and the 4-hydroxyl group of cFuc) are essentially shielded from solvent

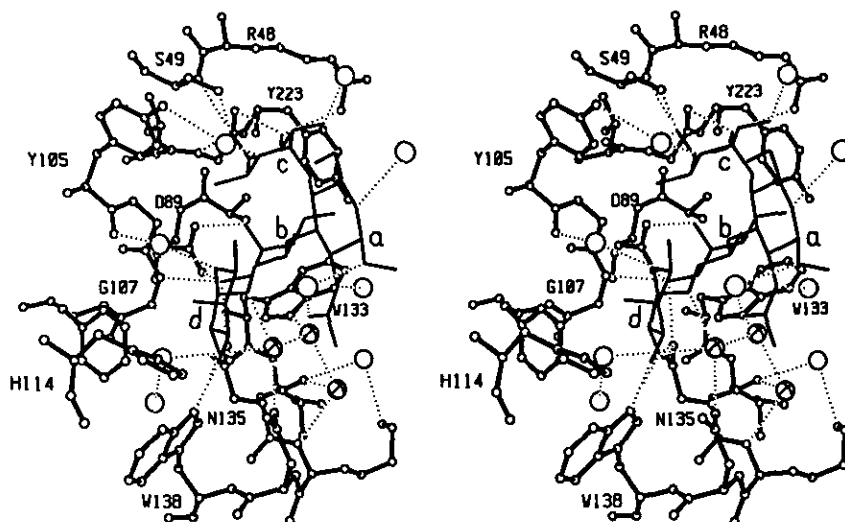


Figure 12. Stereo diagram of the complex of GS4 with Le^b-OME in the area of the carbohydrate-binding site. Non-hydrogen atoms of selected amino acid residues are illustrated as small spheres whereas the Le^b-OME is drawn in stick form. The sugar residues of Le^b-OME are denoted as a, β -D-GlcNAc; b, β -D-Gal; c, α -L-Fuc(1 \rightarrow 4) linked to β -D-GlcNAc; d, α -L-Fuc(1 \rightarrow 2) linked to β -D-Gal. Water molecules are shown as large spheres; the 3 cross-hatched water molecules are conserved in both the native GS4 and the complex structures.

in the complex. The 3-hydroxyl group of **b**Gal has a complete hydrogen-bonding scheme; it forms a hydrogen bond to Asp89 OD2 and accepts hydrogen bonds from Asn135 ND2 and the peptide NH of Gly107. The 4-hydroxyl group of **b**Gal forms a hydrogen bond to Asp89 OD1. The 3- and 4-hydroxyl groups of **b**Gal (which occupy similar positions to those of 2 water molecules in the native structure) are essential for binding, presumably

because they help to neutralize the negative charge on the Asp89 side-chain carboxylate. The occurrence of a *cis*-peptide bond between Ala88 and Asp89 results in the side-chain of Asp89 being approximately at the center of the carbohydrate-binding site (Fig. 12), in the proper orientation to interact with a sugar residue. The third key polar group is the 4-hydroxyl group of **c**Fuc, which is involved in three hydrogen bonds, to the side-chain

Table 6
Hydrogen bonds (<3.3 Å) in the carbohydrate-binding site of the complex

A. Sugar residue	Sugar Atom	Protein atom or water molecule
aGlcNAc	O-1	water458
	N-2	water453
	O-5	water455
	O-7	water458
bGal	O-2	water382
	O-3	Asp89 OD2, Gly107 N, Asn135 ND2
	O-4	Asp89 OD1
	O-2	water456
cFuc	O-3	Arg48 NH2, Gly222 O
	O-4	Ser49 N, Ser49 OG, Gly222 N
	O-5	water457
	O-2	Asn135 ND2, Trp138 NE1, water382, water454
dFuc	O-3	Trp138 NE1
	O-5	water450
B. Protein atom or water	Water molecule	
Tyr105 O	water450	
Tyr105 OH	water457	
Trp133 NE1	water420	
Ile134 O	water409	
Ile134 O	water451	
Lys136 N	water409	
Lys136 NZ	water451	
Asp137 OD1	water382	
water382	water420	
water409	water420	
water452	water454	

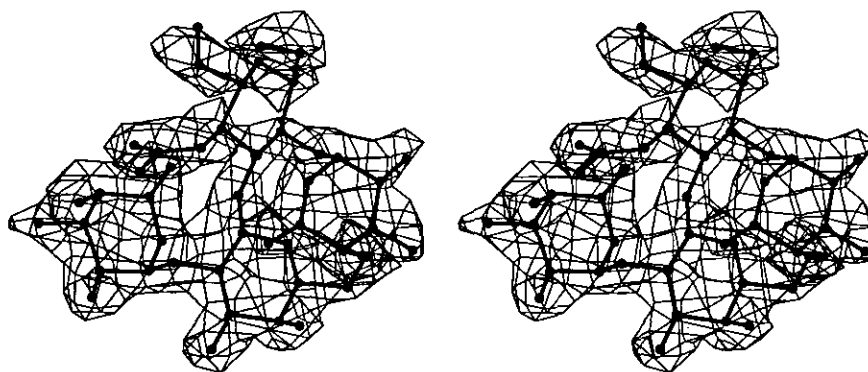


Figure 13. Electron density map in the region of $\text{Le}^b\text{-OMe}$ in the complex. The coefficients for this map are $2|F_o| - |F_c|$ and the phase angles are those calculated from the final model of the complex; the electron density is contoured at 1.0σ . In this view **aGlcNAc** is located at the top centre, **bGal** at the bottom centre, **cFuc** on the right and **dFuc** on the left.

and peptide NH of Ser49 and to the *cis*-peptide NH of Gly222. Thus, the two non-proline *cis*-peptides of GS4, Ala88–Asp89 and Val221–Gly222, are intimately involved in providing a specific carbohydrate-binding site complementary to $\text{Le}^b\text{-OMe}$.

(i) *Peripheral polar interactions*

There are four hydroxyl groups of $\text{Le}^b\text{-OMe}$, located near the periphery of the carbohydrate-binding site in the complex, which form hydrogen bonds with atoms of GS4 and/or with water molecules. These are the 2- and 3-hydroxyl groups of **cFuc** and the 2- and 3-hydroxyl groups of **dFuc** (see Fig. 12). The deoxy-analogs of $\text{Le}^b\text{-OMe}$ which have a hydrogen atom in place of a hydroxyl group at these positions, form complexes with GS4 that have standard free energy changes within the range of -5.5 to -5.7 kcal mol $^{-1}$ (Spohr *et al.*, 1986; Lemieux, 1989). These standard free energy changes are less negative than that found for the $\text{Le}^b\text{-OMe}$ complex, -6.3 kcal mol $^{-1}$ (1 cal = 4.18 J), presumably due to fewer hydrogen bonds. On the other hand, there are three hydroxyl groups of $\text{Le}^b\text{-OMe}$ which are not involved in any hydrogen bonds, the 6-hydroxyl of **aGlcNAc**, the 6-hydroxyl of **bGal** and the 4-hydroxyl of **dFuc**; the corresponding deoxy-

analog complexes have essentially the same standard free energy of formation as that of the $\text{Le}^b\text{-OMe}$ complex (Spohr *et al.*, 1986; Lemieux, 1989).

(j) *Non-polar interactions and atom movements upon complexation*

Table 7 provides a list of 31 distances of $<4 \text{ \AA}$ between non-hydrogen atoms of GS4 and $\text{Le}^b\text{-OMe}$ in the complex. Most of these contacts involve the side-chains of six aromatic amino acids (Tyr105, Phe108, His114, Trp133, Trp138 and Tyr223) which make up a large portion of the carbohydrate-binding site. The side-chain of Trp133 makes more of these contacts than does any other amino acid residue, 12 with the hydrophobic face (and O-6) of **bGal**. The side-chain of Tyr223 also forms contacts with the C-6 and O-6 of **bGal** (as well as with O-3 of **cFuc**). The importance of the non-polar binding interactions of the 6-hydroxymethyl group of **bGal** in complex formation is illustrated by the fact that its replacement by a hydrogen atom produces an inactive analog of $\text{Le}^b\text{-OMe}$ (Lemieux, 1989).

Figure 14 shows the locations of non-hydrogen atoms of the carbohydrate-binding site in the native and the complex. Some atoms of the Trp133 side-

Table 7
van der Waals contacts ($<4 \text{ \AA}$) between GS4 and $\text{Le}^b\text{-OMe}$ in the complex

Sugar residue	Sugar atom	Protein atom
bGal	C-3	Trp133 CD1, NE1
	O-3	Trp133 CD1
	C-4	Trp133 CD1, NE1, CE2
	C-5	Trp133 NE1, CE2, CZ2
	C-6	Trp133 CZ2; Tyr223 CE2, CD2
	O-6	Trp133 CZ2, CH2; Tyr223 CD2, CE2
cFuc	O-2	Arg48 CD
	O-3	Arg48 CD; Tyr223 CB, CD2
	C-6	Tyr105 CD1, CE1
dFuc	C-2	Phe108 CE2
	O-2	Trp138 CD1
	O-3	Trp138 CE2, CZ2
	C-4	His114 ND1
	O-4	Phe108 CE2; His114 CG, ND1, CE1

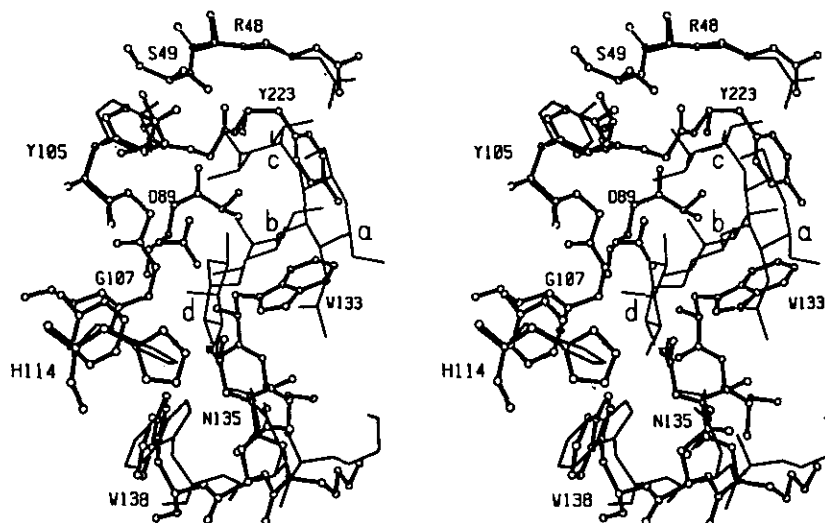


Figure 14. Stereo drawing of the non-hydrogen atoms of the carbohydrate binding site of native GS4 (ball and stick) and the complex (stick bonds).

chain move approximately 0.6 Å from native GS4 positions towards the hydrocarbon face of **bGal** to make the van der Waals contacts described above. The guanidino group of Arg48 moves approximately 0.9 Å from its native GS4 position towards the **Le^b-OMe** to form a hydrogen bond with O-3-**cFuc**; Arg48 CD forms van der Waals contacts with both O-2 and O-3 of **cFuc**. The side-chain of Tyr105 moves approximately 0.7 Å from its native position to avoid steric conflict with **Le^b-OMe** and forms van der Waals contacts with the C-6 methyl group of **cFuc** in the complex. There is a major movement (~2 Å) of the side-chain of His114 from its position in native GS4 to form van der Waals contacts with C-4 and O-4 of **dFuc** and also prevent steric conflict. Movements (~1 to 2 Å) of atoms of the tripeptide Lys136-Asp137-Trp138 allow the formation of the bifurcated hydrogen bond of Trp138 NE1 to both O-2 and O-3 of **dFuc** as well as van der Waals contacts. Thus, there are adjustments of ~1 to 2 Å in positions of certain amino acid residues in the carbohydrate-binding site to optimize the polar and non-polar interactions between GS4 and **Le^b-OMe**. These rearrangements are similar in magnitude (<2 Å) to those found to occur in antibody-antigen complexes (Bhat *et al.*, 1990) and, along with the occurrence of a substantial number of aromatic amino acid residues, may be common aspects of receptor sites in proteins.

The observed glycosidic torsion angles of **Le^b-OMe** in the complex are given in Table 8 along with those obtained by HSEA calculations (Lemieux, 1989). The similarity in these values suggests that the Lewis b tetrasaccharide does not change conformation very much upon binding to GS4. The mean *B*-factor values of relevant residues in the carbohydrate-binding site are given in Table 9 for the complex and native structures. In general, the *B*-factor values of some carbohydrate-binding-site residues are lower in the complex than in native GS4; however, the magnitudes of these decreases are

much smaller than those detected for the complexed and uncomplexed states of wheat germ agglutinin (Wright, 1990).

(k) Carbohydrate-protein complexes

Hydrogen bonding predominates the protein-carbohydrate interactions found in the 1.7 Å resolution structure of the L-arabinose-binding protein and the 1.9 Å structure of the D-galactose-glucose-binding protein (Quiocho, 1990). On the other hand, although there are some important hydrogen bonds, extensive protein-carbohydrate van der Waals contacts (particularly involving aromatic amino acid residues) were found to occur in the 2.05 Å structure of a Fab fragment-dodecasaccharide complex (Cygler *et al.*, 1991) and in the structure of the LOLI-octasaccharide complex (Bourne *et al.*, 1992). The 2.1 Å structure of a lectin-trimannose complex indicated that in addition to hydrogen-bonding between these two moieties, water could also be important in lectin-sugar interactions (Bourne *et al.*, 1990b). Both polar and non-polar interactions are important in the molecular recognition of **Le^b-OMe** by GS4; some water molecules are also associated with the carbohydrate-binding site in the complex.

The 2.0 Å structure of the complex of EcorL with

Table 8

Comparison of ϕ/ψ glycosidic torsion angles (deg.) of **Le^b-OMe** in the crystal structure of the complex with those of HSEA calculations

Linkage	ϕ/ψ (crystal)	ϕ/ψ (HSEA†)
bGal (1 → 3) aGlcNAc	56/18	55/10
cFuc (1 → 4) aGlcNAc	56/29	55/25
dFuc (1 → 2) bGal	53/19	45/15

† Data from Lemieux (1989).

Table 9
Mean B-factor values (\AA^2) in the carbohydrate-binding site of complex (native B-factor values are given in parentheses)

Sugar residue		Protein residue		Solvent
aGlcNAc	42.2	Arg48	42.1 (47.1)	water382 43.6 (36.0)
bGal	30.7	Ser49	28.0 (28.0)	water409 37.7 (40.0)
cFuc	32.7	Asp89	22.5 (20.4)	water420 39.8 (56.7)
dFuc	29.9	Tyr105	32.9 (32.4)	water450 32.7
		Gly106	21.2 (20.7)	water451 63.4
		Gly107	19.9 (24.1)	water452 38.3
		Phe108	27.0 (30.4)	water453 72.1
		Trp133	32.4 (36.6)	water454 37.8
		Ile134	25.5 (25.6)	water455 16.9
		Asn135	29.0 (30.3)	water456 63.1
		Lys136	51.5 (53.6)	water457 60.4
		Asp137	45.6 (49.6)	water458 45.4
		Trp138	38.1 (45.6)	
		Asn139	40.3 (37.6)	
		Gly222	23.8 (26.0)	
		Tyr223	27.8 (31.9)	

lactose (Shaanan *et al.*, 1991) represents a lectin-galactose complex, as does the complex of GS4 with Le^b-OMe. In the EcorL-lactose structure, the 3-hydroxyl group of Gal is hydrogen-bonded to one oxygen atom of the carboxylate group of Asp89, to the peptide NH of Gly107 and to the -NH₂ group of Asn133; this is analogous to the situation in the complex of GS4 with Le^b-OMe. In the EcorL-lactose structure, the 4-hydroxyl group of Gal has two hydrogen bonds, one to the second oxygen atom of the carboxylate of Asp89 and the other to the peptide NH of Ala218, whereas the 4-hydroxyl group of bGal in the complex of GS4 with Le^b-OMe has only the hydrogen bond to Asp89 OD1. The Ala218-Gln219 region of the EcorL carbohydrate-binding site has a different topology than that of glucose mannose-binding lectins; this is presumably

related to the EcorL specificity for galactose (Shaanan *et al.*, 1991). The analogous area in GS4 is Val221-Gly222-Tyr223 which, as a result of the *cis*-peptide bond between residues 221 and 222, has a unique structure in this region. The amino acid residues of the carbohydrate-combining sites of EcorL and GS4 are shown overlapped in Figure 15. It is obvious from an examination of this diagram that EcorL could not bind the Le^b-OMe tetrasaccharide due to a steric conflict that would occur between Ala218-Gln219 of EcorL and cFuc of Le^b-OMe.

This research was supported by grant MA-10162 from the Medical Research Council of Canada and by a NATO Collaborative Research Grant Award no. 890365. Drs U. Spohr and R. U. Lemieux are thanked for providing pure

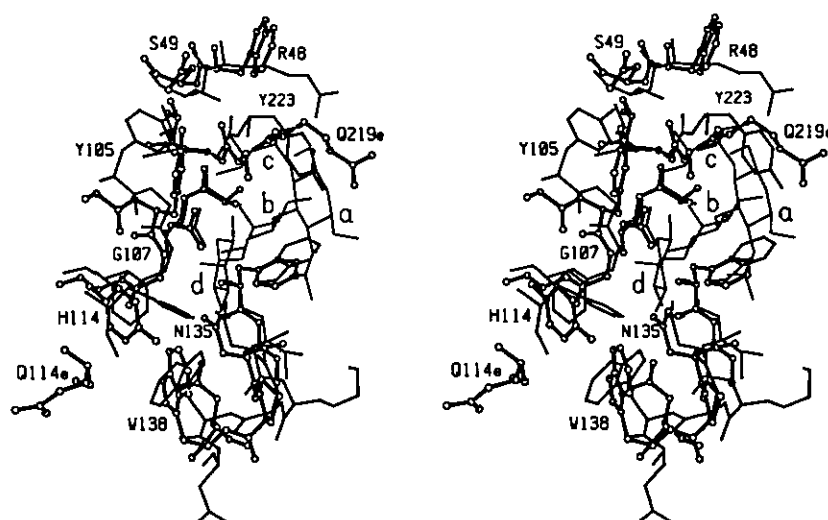


Figure 15. The carbohydrate-binding site of EcorL (ball and stick) and the complex of GS4 with Le^b-OMe (stick bonds) overlapped by use of the program LSQKAB (CCP4, 1979). Selected amino acid residues are labeled. The amino acid sequence numbers of GS4 are given except for Q114e and Q219e which refer to EcorL residues. The Le^b-OMe sugar residues are labeled as a, β -D-GlcNAc; b, β -D-Gal; c, α -L-Fuc(1 \rightarrow 4) linked to β -D-GlcNAc; d, α -L-Fuc(1 \rightarrow 2) linked to β -D-Gal. Note the steric conflict between the Ala218-Gln219 of EcorL and cFuc of Le^b-OMe.

Le^b-OMe, Le^b-OMe analogs and lectin samples as well as for valuable discussions. Dr B. Shaanan is thanked for providing us with the co-ordinates of the EcorL structure.

References

- Becker, J. W., Reeke, G. N., Jr, Wang, J. L., Cunningham, B. A. & Edelman, G. M. (1975). The covalent and three-dimensional structure of concanavalin A. *J. Biol. Chem.* **250**, 1513–1524.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **122**, 535–542.
- Bhat, T. N., Bentley, G. A., Fischman, T. O., Boulot, G. & Poljak, R. J. (1990). Small rearrangements in structures of Fv and Fab fragments of antibody D1.3 on antigen binding. *Nature (London)*, **347**, 483–485.
- Bourne, Y., Abergel, C., Cambillau, C., Frey, M., Rougé, P. & Fontecilla-Camps, J.-C. (1990a). X-ray crystal structure determination and refinement at 1.9 Å resolution of isolectin I from the seeds of *Lathyrus ochrus*. *J. Mol. Biol.* **214**, 571–584.
- Bourne, Y., Rougé, P. & Cambillau, C. (1990b). X-ray structure of a (α-Man(1-3)β-Man(1-4)GlcNAc)-lectin complex at 2.1 Å resolution. *J. Biol. Chem.* **265**, 18161–18165.
- Bourne, Y., Rougé, P. & Cambillau, C. (1992). X-ray structure of a biantennary octasaccharide-lectin complex refined at 2.3 Å resolution. *J. Biol. Chem.* **267**, 197–203.
- Brünger, A. T. (1988). Crystallographic refinement by simulated annealing. Application to a 2.8 Å resolution structure of aspartate aminotransferase. *J. Mol. Biol.* **203**, 803–816.
- CCP4 (1979). *The SERC (U.K.) Collaborative Computing Project No. 4: A Suite of Programs for Protein Crystallography*. Distributed from Daresbury Laboratory, Warrington WA 4AD, U.K.
- Chou, P. Y. & Fasman, G. D. (1977). β-turns in proteins. *J. Mol. Biol.* **115**, 135–175.
- Connolly, M. L. (1983). Analytical molecular surface calculation. *J. Appl. Crystallogr.* **16**, 548–558.
- Cyglér, M., Rose, D. R. & Bundle, D. R. (1991). Recognition of a cell-surface oligosaccharide of pathogenic *Salmonella* by an antibody Fab fragment. *Science*, **253**, 442–445.
- Delbaere, L. T. J., Vandonselaar, M., Prasad, L., Quail, J. W., Nikrad, P. V., Pearlstone, J. R., Carpenter, M. R., Smillie, L. B. & Lemieux, R. U. (1989). Structures of *Griffonia simplicifolia* lectin IV and its complex with a synthetic Lewis b blood group determinant. *Trans. Amer. Crystallogr. Ass.* **25**, 65–76.
- Delbaere, L. T. J., Vandonselaar, M., Prasad, L., Quail, J. W., Pearlstone, J. R., Carpenter, M. R., Smillie, L. B., Nikrad, P. V., Spohr, U. & Lemieux, R. U. (1990). Molecular recognition of a human blood group determinant by a plant lectin. *Can. J. Chem.* **68**, 1116–1121.
- Derewenda, Z., Yariv, J., Helliwell, J. R., Kalb (Gilboa), A. J., Dodson, E. J., Papiz, M. Z., Wan, T. & Campbell, J. (1989). The structure of the saccharide-binding site of concanavalin A. *EMBO J.* **8**, 2189–2193.
- Einspahr, H., Parks, E. H., Suguna, K., Subramanian, E. & Suddath, F. L. (1986). The crystal structure of pea lectin at 3.0-Å resolution. *J. Biol. Chem.* **262**, 16518–16527.
- Evans, S. V. (1992). SETOR: hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graph.* in the press.
- French, S. & Wilson, K. (1978). On the treatment of negative intensity observations. *Acta Crystallogr. Sect. A*, **34**, 517–525.
- Hardman, K. D. & Ainsworth, C. F. (1972). Structure of concanavalin A at 2.4 Å resolution. *Biochemistry*, **11**, 4910–4919.
- Jones, A. T. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* **11**, 268–272.
- Konnert, J. H. & Hendrickson, W. A. (1980). A restrained-parameter thermal-factor refinement procedure. *Acta Crystallogr. Sect. A*, **36**, 344–350.
- Lemieux, R. U. (1989). The origin of the specificity in the recognition of oligosaccharides by proteins. *Chem. Soc. Rev.* **18**, 347–374.
- Lemieux, R. U., Bock, K., Delbaere, L. T. J., Koto, S. & Rao, V. S. (1980). The conformations of oligosaccharides related to the ABH and Lewis human blood group determinants. *Can. J. Chem.* **58**, 631–653.
- Leslie, A. G. W., Brick, P. & Wonacott, A. J. (1986). An improved program package for the measurement of oscillation photographs. *CCP4 Newsletter*, **18**, 33–39.
- Luzzatti, P. V. (1952). Traitement statistique des erreurs dans la détermination des structures cristallines. *Acta Crystallogr.* **5**, 802–810.
- Nikrad, P. V., Pearlstone, J. R., Smillie, L. B. & Lemieux, R. U. (1989). *M_r* differences in *Griffonia simplicifolia* lectin IV subunits are due to oligosaccharide content. In *Abstracts of the Third Symposium of the Protein Society*, July 29–August 2, 1989, S162, Seattle, WA, U.S.A.
- Nikrad, P. V., Pearlstone, J. R., Carpenter, M. R., Lemieux, R. U. & Smillie, L. B. (1990). Molecular-mass heterogeneity of *Griffonia simplicifolia* lectin IV subunits. *Biochem. J.* **272**, 343–350.
- Oatley, S. (1985). *FRODO, UCSD Silicon Graphics IRIS version*, Dept. of Chemistry, University of California at San Diego, CA.
- Quioco, F. A. (1990). Atomic structures of periplasmic binding proteins and the high-affinity active transport systems in bacteria. *Phil. Trans. Roy. Soc. Lond. Ser. B*, **326**, 341–351.
- Ramakrishnan, C. & Ramachandran, G. N. (1965). Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys. J.* **5**, 909–933.
- Reeke, G. N., Jr & Becker, J. W. (1986). Three-dimensional structure of favin: saccharide binding-cyclic permutation in leguminous lectins. *Science*, **234**, 1108–1111.
- Rini, J. M., Hardman, K. D., Einspahr, H. M., Suddath, F. L. & Carver, J. P. (1989). Solution of a pea lectin-trimannoside complex by molecular replacement. *Trans. Amer. Crystallogr. Ass.* **25**, 51–63.
- Roussel, A. & Cambillau, C. (1992). *TURBO-FRODO*, Biographics, Marseilles, France.
- Shaanan, B., Lis, H. & Sharon, N. (1991). Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose. *Science*, **254**, 862–866.
- Shannon, R. D. & Prewitt, C. T. (1969). Ionic radii in oxides and fluorides. *Acta Crystallogr. Sect. B*, **25**, 925–946.
- Shibata, S., Goldstein, I. J. & Baker, D. A. (1982). Isolation and characterization of a Lewis b-active lectin from *Griffonia simplicifolia* seeds. *J. Biol. Chem.* **257**, 9324–9329.

- Shibata, S., Onuma, S. & Inoue, H. (1985). Crystal and molecular structure of trimeric bis(acetylacetonato)manganese(II). *Inorg. Chem.* **24**, 1723-1725.
- Spohr, U. & Lemieux, R. U. (1988). Synthetic conformational and immunochemical studies of modified Lewis b and Y human blood group determinants to serve as probes for the combining site of the lectin IV of *Griffonia simplicifolia*. *Carbohydr. Res.* **174**, 211-237.
- Spohr, U., Back, M. & Lemieux, R. U. (1986). Thermodynamic parameters for the binding of Lewis b tetrasaccharide by the lectin IV of *Griffonia simplicifolia*. In *Abstracts of the XIIIth International Carbohydrate Symposium*, August 10-15, 1986, p. 353, Ithaca, NY.
- Spohr, U., Hindsgaul, O. & Lemieux, R. U. (1985). Molecular recognition. II. The binding of the Lewis b and Y human blood group determinants by the lectin IV of *Griffonia simplicifolia*. *Can. J. Chem.* **63**, 2644-2652.
- Strynadka, N. C. J. & James, M. N. G. (1989). Crystal structures of the helix-loop-helix calcium-binding proteins. *Ann. Rev. Biochem.* **58**, 951-998.
- Vandonselaar, M., Delbaere, L. T. J., Spohr, U. & Lemieux, R. U. (1987). Crystallization of the lectin IV of *Griffonia simplicifolia* and its complexes with the Lewis b and Y human blood determinants. *J. Biol. Chem.* **262**, 10848-10849.
- Vincent, M. G. & Priestle, J. P. (1985). Treatment of intermolecular contacts in the restrained refinement program of Hendrickson & Konnert. *J. Appl. Crystallogr.* **18**, 185-188.
- Wright, C. S. (1990). 2.2 Å resolution structure analysis of two refined *N*-acetylneuraminyllactose-wheat germ agglutinin isolectin complexes. *J. Mol. Biol.* **215**, 635-651.